POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

Corso di Laurea Magistrale in Engineering and Management

Tesi di Laurea di II livello

IL RUOLO DEI DATA BROKER NELL'ECONOMIA DIGITALE

Relatrice: Candidato:

Laura Abrardi Giuseppe Edoardo Pillitteri

Anno Accademico 2023/24

INDICE

Introduzione

Nell'era digitale, l'importanza dei dati è cresciuta esponenzialmente, trasformando il modo in cui le aziende operano e le società interagiscono. Con l'avvento del Big Data e delle tecnologie di analisi avanzata, i dati personali sono diventati una delle risorse più preziose al mondo, superando persino il valore del petrolio.

Negli ultimi dieci anni, dispositivi elettronici, siti web e app hanno sviluppato tecniche sempre più sofisticate per estrarre dati di ogni genere dai consumatori, lo storage dei dati è diventato più economico, spingendo molte aziende a conservare informazioni anche oltre il loro scopo originario. Questo ha incentivato molte piattaforme digitali a fondare i propri modelli di business sulla condivisione di dettagli degli utenti con inserzionisti, aziende private e persino agenzie governative, spesso tramite intermediari noti come data broker.

In questo scenario i data broker, ovvero gli intermediari che raccolgono, aggregano e vendono informazioni personali e comportamentali, hanno assunto un ruolo centrale nelle tecnologie basate sui dati. Essi raccolgono e rivendono enormi quantità di dati sugli individui, comprese informazioni sensibili su dati demografici, finanziari, sanitari, interessi e acquisti. Utilizzano fonti varie, tra cui social media, motori di ricerca, applicazioni, programmi di fedeltà, fornitori di pagamenti e registri pubblici. Ciò che rende i data broker particolarmente potenti è la vastità e la profondità dei loro database, che influenzano molti aspetti della vita quotidiana, dalla pubblicità online alla determinazione dei prezzi, dalla gestione del rischio al punteggio di credito. Tuttavia, diversamente dalle piattaforme digitali più visibili, i data broker non interagiscono direttamente con i consumatori, che spesso non sono neanche consapevoli della loro esistenza.

Il presente lavoro di tesi ha come oggetto l’approfondimento del ruolo dei data broker e le implicazioni di essi nell'economia globale, focalizzandosi su tre aspetti principali: il mercato dei data broker, il loro modello di business e le normative che regolamentano l'uso e la gestione dei dati personali. Attraverso un'analisi dettagliata di queste aree, si intende offrire una visione completa delle opportunità e dei rischi associati all'uso dei dati personali, nonché delle sfide che le aziende e i governi devono affrontare per garantire una corretta gestione e protezione della privacy dei consumatori.

Capitolo 1: Mercato dei Data Broker

1.1 Definizione di Data Broker

I data broker, o intermediari di dati, sono aziende che raccolgono, aggregano e vendono informazioni personali e comportamentali provenienti da diverse fonti, come registri pubblici, transazioni commerciali e attività online. Questo settore è diventato cruciale nel contesto dell'economia digitale, dove i dati costituiscono una risorsa fondamentale per diverse industrie (Sherman, 2021). Questi intermediari, che spesso non hanno una relazione diretta con gli individui di cui raccolgono i dati, acquisiscono tali informazioni da svariate fonti online e offline. Successivamente, vendono queste informazioni a terze parti per vari scopi, inclusi il marketing mirato, la prevenzione delle frodi e la gestione del rischio finanziario. (Christl, 2017).

L'evoluzione dei data broker è strettamente legata alla crescita dell'economia digitale e all'espansione del cosiddetto big data. Il progresso tecnologico, soprattutto l'uso diffuso di Internet e dei social media, ha moltiplicato esponenzialmente la quantità di dati generati ogni giorno. Eric Schmidt sosteneva nel 2010 che la quantità di dati generata in due giorni è pari a tutti i dati della storia umana dagli albori della civiltà fino al 2003 (Arrington, 2010). Le aziende possono utilizzare una parte di questi dati per ottenere informazioni dettagliate sui consumatori, migliorare le loro strategie di marketing e personalizzare le offerte in base ai comportamenti di acquisto (Christl, 2017).

1.2 Big Data

Il concetto di big data è fondamentale per comprendere l'evoluzione dei data broker e il loro ruolo nel contesto dell'economia digitale. I big data si riferiscono a set di dati che sono così vasti, complessi e variegati da non poter essere gestiti con i tradizionali sistemi di gestione dei dati. Secondo Oracle, le caratteristiche principali dei big data possono essere descritte tramite le "cinque V": Volume, Velocità, Varietà, Veridicità e Valore (Oracle, 2024).

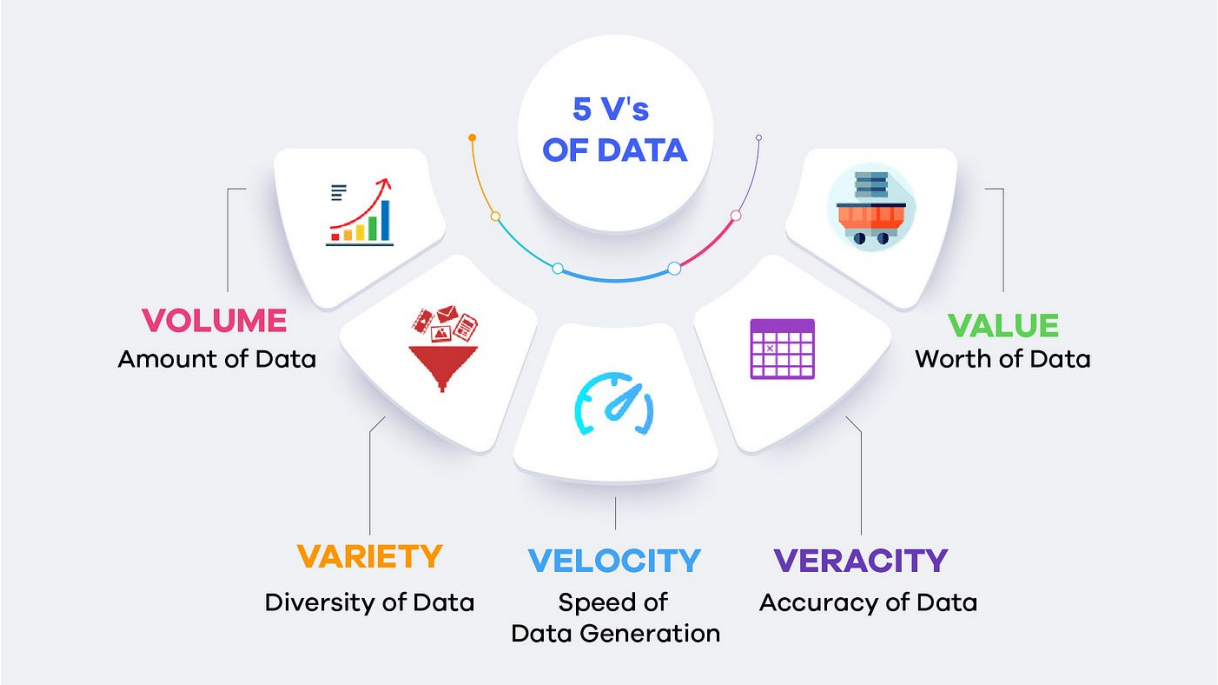

Figura 1: Le 5v dei Big Data.

1. Volume: Questa caratteristica si riferisce alla quantità di dati generati ogni giorno. L'avvento dell'Internet of Things (IoT), dei social media, degli smartphone e delle applicazioni web ha incrementato esponenzialmente il volume di dati prodotti. Ad esempio, si stima che ogni giorno vengano creati miliardi di post sui social media, milioni di transazioni online e terabyte di dati raccolti da sensori e dispositivi connessi (Oracle, 2024). I data broker gestiscono quest'enorme mole di dati personali, questi dati possono essere strutturati, semi-strutturati e non strutturati, e l'attenzione si concentra sulla quantità totale dei dati piuttosto che sul loro contenuto specifico (Excelsior, 2024).

2. Velocità: La velocità si riferisce alla rapidità con cui i dati vengono generati, raccolti, elaborati e analizzati. In molti casi, questi dati vengono prodotti in tempo reale, come avviene nei mercati finanziari, nei sistemi di sorveglianza e nelle transazioni online. La gestione efficace dei big data richiede quindi tecnologie in grado di analizzare e reagire a questi flussi continui di informazioni in tempi brevissimi. La velocità dei dati è quindi diventata una sfida per le aziende, poiché richiede una notevole potenza di elaborazione e infrastrutture che supportino analisi in tempo reale (Excelsior, 2024).

3. Varietà: I dati possono provenire da molteplici fonti e presentarsi in diverse forme, tra cui dati strutturati (come i dati finanziari in tabelle) e dati non strutturati (come

immagini, video, email e post sui social media). La varietà rappresenta una delle sfide più complesse, poiché richiede sistemi di analisi avanzati per estrarre dati di valore da fonti eterogenee. La gestione di questa molteplicità di dati è fondamentale, in quanto i data broker raccolgono informazioni da molteplici e differenti fonti. I dati non strutturati, in particolare, presentano difficoltà, poiché non possono essere facilmente organizzati in database tradizionali di fatti per affrontare questa sfida, le aziende si avvalgono di tecniche di machine learning e intelligenza artificiale che consentono di analizzare e integrare dati provenienti da fonti diverse, offrendo insight più completi (Christl, 2017)(Excelsior, 2024)(Oracle, 2024).

4. Veridicità: La veridicità si riferisce all'affidabilità dei dati, poichè non tutti i dati raccolti sono accurati o pertinenti, e i data broker devono essere in grado di distinguere tra dati validi e informazioni che potrebbero essere errate o fuorvianti. La veridicità è quindi una dimensione critica dei big data, in quanto i dati imprecisi possono portare a decisioni di business sbagliate e a valutazioni del rischio imprecise (Oracle, 2024). Le aziende devono dunque adottare tecnologie e metodologie che permettano di filtrare le informazioni e garantire la qualità dei dati, minimizzando gli errori presenti nei dataset. In sostanza la veridicità non si riferisce soltanto alla raccolta dati ma anche il processo di verifica e di validazione di tali dati (Twetman e Bergmanis-Korats, 2020).

5. Valore: Il valore si riferisce all'utilità dei dati raccolti, i big data diventano davvero preziosi solo quando vengono analizzati per ottenere insight che aiutano le aziende a prendere decisioni informate. I data broker sono specializzati nel trasformare grandi quantità di dati grezzi in informazioni utili per il marketing, la gestione del rischio e la prevenzione delle frodi. Per i data broker, il valore risiede nella loro capacità di convertire grandi quantità di dati grezzi in informazioni utilizzabili per migliorare le strategie di marketing, la gestione dei rischi e la customer experience (Kaspersky, 2024). Questo permette alle aziende di sfruttare i dati in modo da ottenere vantaggi competitivi e migliorare le loro operazioni commerciali. Il valore è l'elemento che giustifica gli investimenti nei big data e nelle tecnologie di analisi avanzata, trasformando i dati in uno strumento strategico per guidare il business verso il successo (Oracle, 2024) (Twetman e Bergmanis-Korats, 2020).

I big data rappresentano una risorsa essenziale per i data broker, che sfruttano le "cinque V" per raccogliere, analizzare e vendere informazioni dettagliate sui consumatori. Grazie ai progressi nel cloud computing, nell'intelligenza artificiale e nelle tecnologie di analisi dei dati, i broker

possono ora processare enormi volumi di dati in tempo reale, fornendo insight dettagliati sui comportamenti e le preferenze dei consumatori (Oracle, 2024)(Twetman e Bergmanis-Korats, 2020). La varietà delle informazioni raccolte permette ai data broker di creare profili dei consumatori estremamente dettagliati, che includono dati demografici, comportamentali, finanziari e persino psicografici. Questi profili sono quindi venduti a terzi, come aziende di marketing, istituzioni finanziarie e agenzie governative, per vari scopi, tra cui il marketing mirato, la gestione del rischio e la prevenzione delle frodi (Christl, 2017).

1.3 Evoluzione Storica dei Data Broker

L'evoluzione dei data broker è strettamente legata alla trasformazione dell'economia digitale e all'uso massiccio dei dati personali da parte delle aziende. L'idea di raccogliere e vendere informazioni personali non è recente; le prime forme di intermediazione dei dati risalgono al XX secolo, quando le aziende iniziarono a raccogliere dati demografici e comportamentali dai registri pubblici, dalle operazioni finanziarie e da altre fonti offline. Tuttavia, il vero sviluppo del settore si è avuto con l'avvento di Internet e delle nuove tecnologie digitali.

Negli anni '70 e '80, alcune aziende come Equifax e Acxiom cominciarono a raccogliere dati sui consumatori tramite registri pubblici e transazioni commerciali, con l'obiettivo di fornire rapporti finanziari e informazioni sui consumatori a imprese e istituzioni finanziarie. Il settore era ancora limitato, con un focus principalmente su dati finanziari e relativi alla solvibilità, tuttavia, durante gli anni '90, con la crescita di Internet e del commercio elettronico, i data broker come Acxiom e Experian si sono trasformati in giganti globali della raccolta di dati (WebFX, 2024). Ad esempio, Acxiom nel 2012 dichiarava di avere circa 500 milioni di profili di consumatori, mentre nel 2023 ha raggiunto i 2,5 miliardi di persone con oltre 11.000 punti dati per individuo (Kemp, 2024).

Questa crescita esponenziale è stata alimentata dalla crescente disponibilità di dati digitali e dall’uso dei primi cookie, introdotti per tracciare le abitudini di navigazione degli utenti online. Le informazioni raccolte venivano aggregate per creare profili dettagliati, utilizzati da aziende per scopi di marketing, targetizzazione pubblicitaria e gestione del rischio. Questo ha segnato l'inizio di una nuova era in cui i dati digitali sono diventati la merce principale di scambio (History Tools, 2024).

A partire dai primi anni 2000, l'esplosione del Big Data ha ulteriormente trasformato il settore dei data broker. Con l'avvento dei social media, dei dispositivi mobili e delle tecnologie di tracciamento più avanzate come il fingerprinting dei dispositivi e i pixel di tracciamento, i data

broker hanno ampliato la loro capacità di raccogliere informazioni in tempo reale da milioni di utenti in tutto il mondo. Aziende come Google e Facebook hanno iniziato a raccogliere dati direttamente dagli utenti, mentre i data broker tradizionali si sono concentrati sulla raccolta da fonti terze, come gli acquisti online e le app mobili, per creare "super profili" completi e precisi (Moes, 2023).

Durante questo periodo, i data broker hanno cominciato a raccogliere tutti questi dati per essere usati per la pubblicità mirata. Utilizzando tecniche di machine learning e intelligenza artificiale, i broker segmentano i dati in base a caratteristiche demografiche, comportamentali e di localizzazione, consentendo alle aziende di indirizzare gli annunci pubblicitari con una precisione mai vista prima. Questi profili vengono poi venduti a inserzionisti, aziende di marketing e altre entità commerciali.

Con la crescita dell'intelligenza artificiale e l'uso di tecnologie avanzate come la blockchain per la protezione dei dati, il settore dei data broker è destinato a evolversi ulteriormente, tuttavia, le controversie etiche e legali legate alla raccolta e all'uso dei dati continueranno a essere presenti. L'industria dei data broker si trova di fronte a un futuro incerto, in cui la regolamentazione e la crescente consapevolezza dei consumatori potrebbero limitare le sue attività, o costringerla a trovare nuovi modi per operare all'interno dei confini normativi.

1.4 Tipologie di Data Broker

I data broker possono essere classificati secondo due criteri principali: il metodo di acquisizione dei dati e la tipologia di dati che raccolgono e rivendono. Questa duplice classificazione offre una visione più chiara delle diverse attività svolte dai data broker nell'attuale economia digitale.

Nella prima modalità di classificazione analizzata si distinguono i first-party data broker, che raccolgono dati direttamente dai consumatori attraverso una relazione diretta (come transazioni su un sito web o programmi di fedeltà), e i third-party data broker, che invece raccolgono informazioni da varie fonti esterne, spesso senza una relazione diretta con gli utenti, come registri pubblici, social media e partner commerciali.

Nella seconda modalità di classificazione analizzata si distinguono i data broker che raccolgono diverse categorie di dati, come dati finanziari, dati per la mitigazione dei rischi, dati per il marketing e la pubblicità, dati personali (people search), e dati sanitari. Questa categorizzazione offre una panoramica dei servizi che i data broker possono offrire alle aziende e alle istituzioni in base alle informazioni specifiche che sono in grado di raccogliere, analizzare e rivendere.

1.4.1 Classificazione basata sul metodo di raccolta dati

Secondo il documento dell'Australian Competition and Consumer Commission (ACCC, 2024), i data broker possono essere suddivisi principalmente in due categorie: first-party data broker e third-party data broker. Questa distinzione è importante perché evidenzia la fonte dei dati raccolti e il modo in cui vengono utilizzati e condivisi sul mercato.

1.4.1.1 First-Party Data Broker

I first-party data broker sono aziende che raccolgono informazioni direttamente dai propri consumatori ciò significa che ottengono i dati tramite le interazioni dirette con i clienti, come gli acquisti effettuati su un sito web, la partecipazione a programmi di fedeltà, o attraverso l'uso di applicazioni aziendali. Una volta raccolte queste informazioni, i first-party data broker possono venderle o condividerle con altre aziende, come ad esempio società di marketing o istituzioni finanziarie. Questi tipi di broker sono spesso più trasparenti nella raccolta e gestione dei dati poiché, avendo una relazione diretta con i consumatori, sono in grado di ottenere il loro consenso esplicito per l'uso delle informazioni. Tuttavia, anche in questo caso, l'utente potrebbe non essere pienamente consapevole dell'entità dei dati raccolti e della loro successiva distribuzione a terzi (ACCC, 2024).

1.4.1.2 Third-Party Data Broker

I third-party data broker sono aziende che raccolgono dati su individui attraverso una vasta gamma di fonti di terze parti, senza avere una relazione diretta con i consumatori stessi. Queste fonti possono includere:

- Piattaforme digitali (come social media e siti web).
- Web scraping e l'uso di cookie.
- Applicazioni mobili tramite l'utilizzo di kit di sviluppo software (SDK).
- Governo e altre organizzazioni come database pubblici, registri elettorali e progetti di open data.
- Schemi di fidelizzazione e altri data broker che vendono o scambiano dati tra loro.

I third-party data broker spesso aggiungono valore ai dati raccolti attraverso tecniche di analisi e aggregazione avanzate, creando così prodotti e servizi personalizzati che possono essere venduti o concessi in licenza alle aziende. Questa tipologia di data broker è più problematica in termini di trasparenza e protezione della privacy, poiché i consumatori sono spesso all’oscuro

della raccolta e dell'uso dei loro dati da parte di terze parti. Di conseguenza, i consumatori raramente danno il consenso esplicito per la raccolta e l'analisi delle loro informazioni da parte di questi broker.

La differenza tra queste due categorie di data broker sottolinea l'importanza di una maggiore trasparenza e regolamentazione nell'industria della gestione dei dati, poiché le pratiche dei third-party data broker sollevano più preoccupazioni in termini di privacy e controllo dei dati da parte degli utenti (ACCC, 2024).

1.4.2 Classificazione basata sulla tipologia di dati

In questa classificazione vedremo i data broker suddivisi in diverse categorie, a seconda del tipo di informazioni che raccolgono e degli scopi per cui tali dati vengono utilizzati. Questa classificazione aiuta a comprendere meglio la complessità e la varietà di operazioni che queste aziende conducono nel mercato dei dati.

1.4.2.1 Data Broker per il Marketing e la Pubblicità

Uno dei tipi più comuni di data broker è quello che si occupa della raccolta e vendita di informazioni per scopi di marketing e pubblicità, raccogliendo informazioni sui consumatori per aiutare le aziende a targetizzare il loro pubblico in modo più preciso. Questi broker raccolgono dati demografici e comportamentali, come età, sesso, reddito, interessi e cronologia di acquisto, per creare profili dettagliati che vengono utilizzati dalle aziende per creare campagne pubblicitarie personalizzate (SoftwareLab, 2024). Inoltre, offrono servizi di "append" ovvero aggiungono ulteriori informazioni ai profili dei consumatori esistenti, come indirizzo o cronologia degli acquisti. Acxiom e Epsilon e Oracle sono i principali leader di data broker in questo settore, mentre un nuovo gruppo emergente di broker si è specializzato nella fornitura di dati di localizzazione per inserzionisti e marketer. Un esempio è Veraset, che mette a disposizione dati GPS raccolti da migliaia di applicazioni, permettendo un monitoraggio accurato delle abitudini di movimento delle persone (Kemp, 2022).

1.4.2.2 Data Broker Finanziari e per la Gestione del Rischio

Un altro settore cruciale in cui operano i data broker è la prevenzione delle frodi e la gestione del rischio. In questo ambito, i broker forniscono dati a istituti finanziari, compagnie assicurative e altre organizzazioni che devono valutare i rischi associati a clienti o transazioni. Ad esempio, raccolgono e analizzano informazioni sulla storia creditizia di un individuo, sui suoi comportamenti di spesa e sul rischio di insolvenza, permettendo alle aziende di identificare potenziali frodi o rischi finanziari. Queste informazioni possono includere la cronologia

lavorativa, il livello salariale, la verifica delle licenze professionali e altri dati finanziari. Le istituzioni utilizzano questi dati per diversi scopi, come verificare che le informazioni fornite da un consumatore in una richiesta di prestito corrispondano a quelle presenti nei database dei data broker, o per calcolare la probabilità che un cliente possa non restituire un prestito (Kaspersky, 2024). I data broker specializzati in questo settore collaborano frequentemente con le istituzioni finanziarie per confermare l'identità dei clienti e prevenire crimini finanziari, come il furto d'identità. Tra i principali attori in questo campo troviamo LexisNexis Risk Solutions e CoreLogic, che offrono servizi di verifica dell'identità e gestione del rischio attraverso l'uso di vasti database contenenti sia dati pubblici che privati (Kemp, 2022).

1.4.2.3 Data Broker per la Ricerca di Persone

Esistono broker specializzati nella ricerca di persone, come Spokeo, PeopleFinder, e White Pages, che raccolgono e aggregano informazioni personali da registri pubblici, social media e fonti commerciali. Questi broker creano profili dettagliati, molto più completi di un semplice elenco telefonico, facilitando la ricerca di un individuo. I loro portali permettono agli utenti di accedere a informazioni approfondite su altre persone, come nomi, indirizzi, numeri di telefono, stato civile e persino dettagli sulla storia lavorativa. Alcuni di questi siti offrono servizi di "background check" completi, che possono includere registri di arresto, cronologia lavorativa e altre informazioni personali. Questo tipo di data broker viene spesso utilizzato in vari contesti, tra cui il controllo dei precedenti per le assunzioni, le verifiche per l’affitto di proprietà o le ricerche genealogiche. Tuttavia, la raccolta e l’uso di tali informazioni sollevano importanti questioni etiche, soprattutto quando si tratta di possibili violazioni della privacy (Kemp, 2022).

1.4.2.4 Data Broker di Dati Sanitari

I data broker sanitari raccolgono e aggregano informazioni dettagliate sulle condizioni mediche delle persone, inclusi i farmaci acquistati e i sintomi cercati online. Questi dati vengono utilizzati per creare profili completi che possono contenere informazioni su patologie specifiche, come depressione, ansia e altre condizioni di salute, questi profili vengono poi rivenduti a una varietà di clienti, tra cui compagnie di assicurazione sanitaria, aziende farmaceutiche e marketer. Le assicurazioni, in particolare, possono utilizzare queste informazioni per determinare le tariffe da applicare ai consumatori in base al loro profilo sanitario, mentre le aziende farmaceutiche le impiegano per attività di marketing mirato. Queste pratiche sollevano importanti questioni etiche e di privacy, poiché molte persone non sono consapevoli del fatto che i loro dati sanitari vengano raccolti e condivisi con terze parti senza il loro consenso esplicito. Inoltre, poiché i data broker sanitari spesso non sono soggetti a

normative sulla privacy, come l'Health Insurance Portability and Accountability Act (HIPAA) negli Stati Uniti, possono vendere questi dati sensibili senza particolari restrizioni e senza un controllo rigoroso. Questa mancanza di regolamentazione permette ai broker di segmentare i consumatori in base alle loro condizioni di salute, fornendo alle aziende profili dettagliati per finalità di marketing e valutazione del rischio (Kemp, 2022) (McCormack, 2023).

1.5 Disparità di Informazione

"The Rich See a Different Internet Than the Poor", con questa frase, Michael Fertik, nel 2013, in un suo articolo, descriveva un mondo digitale sempre più manipolato da terze parti, dove algoritmi e raccolta dati stabiliscono cosa vediamo, acquistiamo e con chi interagiamo (Fertik, 2013). La "personalizzazione" apparentemente offre libertà di scelta, ma in realtà nasconde una forma di discriminazione, trasformando Internet in uno spazio differenziato in base all'identità dell'utente: ricchi e poveri vivono esperienze online radicalmente diverse. Questo fenomeno è alimentato dalla vendita di dati personali da parte dei data broker poiché i dati inizialmente raccolti per scopi di marketing sono oggi sfruttati da compagnie assicurative e mediche per decidere come interagire con gli utenti, spesso senza che essi ne siano consapevoli. Di conseguenza, il 99% degli utenti vive "dall’altra parte di uno specchio unidirezionale", dove una piccola élite manipola la loro esperienza online.

Questa dinamica può portare a vere e proprie discriminazioni, come nell’accesso al credito, le istituzioni finanziarie possono evitare di fare offerte a persone considerate "meno attraenti" dal punto di vista creditizio, aggirando di fatto le leggi che vietano la discriminazione. Anche i prezzi online sono spesso personalizzati in base alle abitudini dell'utente, alla sua posizione geografica e al momento in cui visita il sito, generando situazioni in cui alcuni pagano di più per lo stesso prodotto senza saperlo. Questa segmentazione non riguarda solo il commercio, ma si estende anche alla vita politica digitale: gli algoritmi creano "camere d'eco" che rafforzano le opinioni esistenti, contribuendo alla polarizzazione politica.

Uno studio recente intitolato *"Measuring Biases in a Data Broker's Coverage"* (Kaplan et al., 2022) ha ulteriormente evidenziato come questa personalizzazione digitale aumenti il divario sociale. La ricerca, condotta negli Stati Uniti, mostra che l'assenza di un sistema di identificazione standard spinge aziende e organizzazioni a utilizzare i data broker per verificare l’identità dei clienti e condurre controlli di background. Decisioni importanti, come quelle riguardanti l’alloggio, il credito e il lavoro si basano spesso sulle informazioni fornite da questi broker, che operano però con poca trasparenza.

Lo studio si è concentrato su Experian, uno dei principali data broker, e ha analizzato la sua copertura nella Carolina del Nord confrontando i dati provenienti dal suo database con i registri elettorali della Carolina del Nord e i dati del censimento statunitense. I risultati mostrano che la rappresentazione nei database di Experian è incompleta, soprattutto per i gruppi storicamente svantaggiati: persone più giovani, minoranze etniche e chi vive in aree a basso reddito. Anche quando sono presenti, i loro dati risultano spesso imprecisi rispetto a quelli dei bianchi e dei residenti in aree più ricche, come ad esempio, l'identificazione degli elettori ispanici che ha il 50% di probabilità in più di non essere corretta rispetto a quella degli elettori bianchi.

Queste disparità possono amplificare ulteriormente le disuguaglianze sociali, rendendo l'accesso a opportunità fondamentali ancora più difficile per le popolazioni vulnerabili. Gli autori dello studio sottolineano la necessità di maggiore trasparenza e di meccanismi più semplici per correggere i dati errati e suggeriscono che le aziende che si affidano ai data broker per decisioni critiche dovrebbero prestare particolare attenzione ai casi in cui l'identità non può essere verificata, cercando di offrire metodi alternativi senza penalizzare gli individui.

1.6 Principali Attori del Mercato dei Data Broker

Il mercato dei data broker è dominato da un gruppo di aziende multinazionali che gestiscono enormi quantità di informazioni personali e dati sensibili raccolti da fonti online e offline. In questo paragrafo esamineremo i principali attori globali del settore e analizzeremo le loro attività, le tecnologie che utilizzano e l'impatto che hanno sul mercato. Ogni azienda ha sviluppato un modello di business unico, ma tutte condividono un obiettivo comune: raccogliere, analizzare e vendere dati per scopi commerciali.

1.6.1 Acxiom

Figura 2: Logo Acxiom.

Acxiom LLC è una delle più grandi e influenti aziende operanti nel settore dei data broker. Fondata nel 1969 come Demographics Inc., ha rapidamente evoluto il proprio modello di

business passando dalla fornitura di dati demografici alla raccolta, aggregazione e vendita di dati personali a livello globale. Ad oggi, Acxiom ha trasformato il mercato del brokeraggio dei dati, diventando un attore chiave nella raccolta e gestione di enormi volumi di dati su individui e consumatori di tutto il mondo. La sede principale di Acxiom si trova a Conway, Arkansas, dove fu fondata, ma l'azienda ha una presenza globale, con uffici in Nord America, Europa, Asia e America Latina. Questo permette ad Acxiom di operare su scala mondiale, raccogliendo dati da una vasta gamma di fonti, inclusi registri pubblici, transazioni commerciali, acquisti online e attività sui social media.

Nel 2018, Acxiom ha venduto la sua divisione marketing a Interpublic Group per 2,3 miliardi di dollari, concentrandosi sulla piattaforma LiveRamp, una delle più avanzate piattaforme di gestione dei dati (Data Management Platform - DMP). Questa piattaforma consente alle aziende di utilizzare dati personali in modo anonimo per scopi di marketing e pubblicità mirata, mantenendo una forte leadership nel settore dei big data e delle soluzioni di marketing basate sui dati.

Acxiom opera principalmente come data broker e data integrator, raccogliendo, elaborando e vendendo dati su individui a una vasta gamma di clienti. Questi dati vengono utilizzati per segmentare il mercato, migliorare le campagne di marketing, personalizzare le offerte e ottimizzare la customer experience. La piattaforma LiveRamp è particolarmente importante, poiché permette di collegare i dati di diverse fonti per creare profili completi dei consumatori, che possono essere utilizzati per targetizzare le campagne pubblicitarie in modo estremamente preciso.

I dati trattati da Acxiom includono:

- Dati demografici: informazioni di base come età, sesso, reddito, livello di istruzione, e composizione familiare.
- Dati comportamentali: abitudini di acquisto, cronologia di navigazione web, interazioni sui social media.
- Dati geografici: informazioni sulla localizzazione ottenute tramite indirizzi IP, GPS e altre fonti.
- Dati transazionali: acquisti effettuati online e offline, dati delle carte di credito e programmi di fidelizzazione.

Uno dei vantaggi competitivi di Acxiom è la sua capacità di combinare dati strutturati e non strutturati provenienti da diverse fonti, utilizzando avanzate tecniche di machine learning e big data per analizzare enormi set di dati e fornire insight dettagliati alle aziende. Questi insight permettono alle aziende clienti di prevedere comportamenti di acquisto, migliorare la fidelizzazione dei clienti e massimizzare il ritorno sugli investimenti delle loro campagne di marketing

Acxiom serve una vasta gamma di settori, inclusi:

- Retail: supporta i rivenditori nell'ottimizzazione delle loro strategie di marketing.
- Telecomunicazioni: aiuta le aziende a personalizzare le offerte per i clienti in base ai dati comportamentali.
- Finanza: fornisce dati per valutazioni del rischio e decisioni di credito.
- Salute: supporta aziende sanitarie con soluzioni basate sui dati per migliorare l'engagement dei pazienti e la personalizzazione delle cure.

L'azienda vanta clienti di alto profilo come Procter & Gamble, Unilever, e AT&T, e collabora con oltre 7.000 aziende a livello globale.

Acxiom è un attore dominante in un settore che vale centinaia di miliardi di dollari all'anno. La società gestisce oltre 3.000 punti dati per ogni individuo nei suoi archivi e possiede dati su circa 2,5 miliardi di consumatori in tutto il mondo. La sua capacità di raccogliere, analizzare e vendere dati su tale scala le consente di generare entrate significative, contribuendo in modo cruciale al mercato globale dei big data. Acxiom, come molti altri data broker, è stato al centro di controversie legate alla privacy e alla protezione dei dati. Le preoccupazioni principali riguardano la mancanza di trasparenza sulla raccolta e l'uso dei dati da parte dell'azienda. Acxiom fornisce meccanismi di opt-out che consentono ai consumatori di chiedere la rimozione dei propri dati dai database dell’azienda; tuttavia, questi strumenti sono difficili da utilizzare e anche poco conosciuti.

Inoltre, con l'introduzione di normative più rigorose sulla privacy dei dati, come il GDPR in Europa e il CCPA in California, Acxiom ha dovuto adattare le proprie operazioni per conformarsi a questi nuovi standard. Tuttavia, resta ancora molto da fare per garantire che i consumatori abbiano un maggiore controllo sui propri dati personali e che vengano ridotti i rischi di data breach e di uso improprio dei dati (Acxiom, 2024)(Dhaliwal, 2024)(Sherman, 2021).

1.6.2 Experian

Figura 3: Logo Experian.

Experian è una delle tre principali agenzie di credito (credit bureau) a livello globale, insieme a Equifax e TransUnion. Fondata nel 1996, Experian è diventata una delle più grandi aziende nel settore dei dati finanziari e della gestione del credito. L'azienda, che ha operazioni in oltre 44 paesi e sede centrale a Dublino, serve milioni di consumatori e imprese in tutto il mondo. La sua missione principale è aiutare le aziende e le istituzioni finanziarie a prendere decisioni più informate sui clienti attraverso l'accesso a dati e informazioni creditizie. Experian, oltre alla gestione dei dati creditizi, è anche un attore importante nel settore del brokeraggio dei dati, vendendo informazioni personali a terzi per scopi di marketing e valutazione del rischio finanziario.

Nel 2023, Experian ha generato un fatturato globale di circa 5 miliardi di dollari grazie ai suoi servizi, consolidandosi come uno degli attori chiave nel mercato dei dati. Le sue operazioni comprendono sia la gestione dei dati finanziari che la fornitura di soluzioni di marketing basate sui dati, oltre a servizi di protezione contro le frodi e analisi avanzata dei rischi.

Experian si concentra principalmente su tre settori: servizi di credito, decision analytics, e marketing services.

- Servizi di credito: Experian raccoglie e gestisce informazioni finanziarie su milioni di individui e aziende, inclusi i punteggi di credito, la cronologia dei pagamenti, i crediti in corso e altre informazioni utili per la valutazione della solvibilità dei consumatori. Questi dati sono utilizzati principalmente da banche e istituzioni finanziarie per prendere decisioni riguardanti l'erogazione di prestiti, l'approvazione di carte di credito e altre operazioni finanziarie.

- Decision Analytics: Experian utilizza tecnologie di machine learning e analisi predittiva per valutare i rischi associati a operazioni finanziarie o commerciali. Questo settore consente ai clienti di Experian di prevedere comportamenti di credito e di acquisto, di individuare potenziali frodi e di ottimizzare le decisioni relative ai clienti. Il suo software di analisi è in grado di integrare dati da più fonti, migliorando la capacità delle aziende di prendere decisioni informate sui consumatori.
- Marketing Services: Un'altra area significativa di attività per Experian è il supporto alle aziende nell'ottimizzazione delle campagne di marketing. Experian raccoglie dati demografici, comportamentali e di localizzazione per aiutare le aziende a individuare meglio i loro clienti e migliorare il rendimento delle campagne pubblicitarie. Questo settore è particolarmente importante per i rivenditori e le aziende che desiderano migliorare l'efficienza delle loro strategie di marketing digitale.

Experian serve principalmente il settore finanziario, lavorando con banche, istituti di credito, compagnie assicurative e altri attori del mondo delle finanze per aiutarli a prendere decisioni basate su dati affidabili. Inoltre, l'azienda collabora con il settore del retail, delle telecomunicazioni e della sanità, offrendo soluzioni che migliorano la gestione dei dati dei clienti e la prevenzione delle frodi.

Uno dei principali settori in cui Experian ha fatto la differenza è quello dei crediti al consumo: attraverso la gestione di database contenenti informazioni sui punteggi di credito di milioni di consumatori, Experian supporta le banche e altre istituzioni finanziarie nel determinare se concedere prestiti o linee di credito a determinati clienti. Questo servizio ha un impatto diretto sulle decisioni di credito di milioni di persone in tutto il mondo.

Con una presenza in oltre 44 paesi e un database che contiene informazioni su circa 1 miliardo di individui e 145 milioni di aziende in tutto il mondo, Experian è uno dei più grandi detentori di dati finanziari e personali a livello globale. L'azienda continua a crescere rapidamente, espandendo le proprie operazioni in nuovi mercati e investendo in tecnologie avanzate come l'intelligenza artificiale e il machine learning per migliorare la qualità dei suoi servizi e dei dati che gestisce.

Il mercato dei servizi di credito è una delle principali fonti di reddito per Experian. In particolare, il mercato nordamericano rappresenta la fetta più grande del fatturato dell'azienda, seguito dai mercati europei e asiatici. L'espansione nei mercati emergenti, come l'America Latina e l'Asia, è uno dei focus principali di Experian, che continua a cercare nuove opportunità

di crescita nei paesi in via di sviluppo. Experian fu oggetto di critiche e controversie poiché fu al centro di uno degli eventi più significativi in termini di sicurezza e privacy, ovvero il data breach del 2015, in cui gli hacker hanno compromesso i dati di 15 milioni di utenti. Questo ha portato a un'ondata di critiche riguardo alla gestione dei dati e ha sollevato dubbi sulle capacità dell'azienda di proteggere le informazioni sensibili dei consumatori.

Negli anni successivi, Experian ha lavorato per migliorare la sicurezza dei suoi database e conformarsi alle normative internazionali sulla privacy, come il GDPR in Europa e il CCPA in California. Tuttavia, rimangono dubbi per quanto riguarda la protezione della privacy dei dati e l'uso etico delle informazioni che raccoglie. (Twingate, 2024)(Experian, 2024)

1.6.3 Equifax

Figura 4: Logo Equifax.

Equifax, fondata nel 1899, è una delle tre principali agenzie di credito al mondo, insieme a Experian e TransUnion. Con sede centrale ad Atlanta, Georgia, Equifax opera in oltre 24 paesi e impiega più di 11.000 dipendenti a livello globale. L'azienda si distingue per la sua capacità di fornire soluzioni di data analytics, tecnologie di cloud computing, e servizi finanziari, aiutando banche, istituzioni governative e aziende a prendere decisioni critiche basate su dati accurati e aggiornati.

Equifax fornisce informazioni creditizie e finanziarie che consentono ai suoi clienti di valutare la solvibilità dei consumatori, gestire i rischi finanziari e personalizzare le offerte di credito. L'azienda è responsabile della raccolta e dell'elaborazione di enormi volumi di dati personali e finanziari, che sono poi utilizzati per fornire servizi come punteggi di credito, rapporti sul credito e protezione contro il furto d'identità.

Nel 2023, Equifax ha generato circa 5,3 miliardi di dollari di entrate, con una crescita significativa guidata dall'espansione delle sue soluzioni cloud e dall'aumento della domanda di servizi di verifica delle identità e di protezione contro le frodi.

Equifax opera attraverso tre principali segmenti:

1. U.S. Information Solutions (USIS): Questo segmento fornisce informazioni sui consumatori e soluzioni commerciali negli Stati Uniti, includendo servizi di analisi finanziaria, gestione del rischio e frodi, nonché report sui mutui e sui crediti.
2. Workforce Solutions: Questo segmento offre servizi di verifica dell'occupazione e del reddito, nonché soluzioni per la gestione delle risorse umane e il rispetto delle normative, inclusa la gestione delle tasse e dei benefit.
3. International: Questo segmento copre le operazioni al di fuori degli Stati Uniti, tra cui Europa, America Latina e Asia-Pacifico, fornendo prodotti simili adattati alle normative locali.

Un aspetto chiave delle operazioni di Equifax è la sua piattaforma cloud nativa, introdotta negli ultimi anni. Equifax ha investito oltre 1,5 miliardi di dollari nella trasformazione digitale e nella migrazione di oltre 50 miliardi di record su questa piattaforma. Questo permette una maggiore efficienza nel trattamento dei dati, riducendo i tempi di elaborazione e migliorando la protezione dei dati.

Equifax serve una vasta gamma di settori, tra cui:

- Istituzioni finanziarie: Fornendo punteggi di credito e rapporti finanziari che aiutano a valutare la solvibilità dei consumatori e a gestire il rischio associato ai prestiti.
- Compagnie assicurative: Offrendo soluzioni di verifica dell'identità e valutazione del rischio per prevenire frodi e migliorare l'accuratezza delle polizze assicurative.
- Governi e agenzie governative: Equifax supporta le agenzie governative nella verifica dell'idoneità dei benefici sociali e nella gestione dei crediti fiscali.

Nel 2023, Equifax ha registrato una crescita del 10% nelle sue entrate non legate ai mutui, dimostrando la diversificazione della sua base clienti. I suoi servizi di verifica dell'occupazione e del reddito hanno avuto una crescita significativa, con un incremento del 11% nelle entrate derivanti da tali soluzioni. Inoltre, la società ha investito in acquisizioni strategiche per espandere i propri asset di dati e migliorare le capacità analitiche.

Equifax ha affrontato una delle più grandi violazioni di dati della storia nel 2017, quando un attacco informatico ha esposto le informazioni personali di 147 milioni di individui, inclusi numeri di previdenza sociale e dati finanziari sensibili. Questo scandalo ha portato a numerose critiche sull'inadeguata sicurezza dei dati e ha spinto l'azienda a investire pesantemente in miglioramenti della sicurezza informatica e della protezione dei dati.

L'azienda ha risposto con un piano di trasformazione a lungo termine, che includeva una completa ristrutturazione della sua infrastruttura IT e investimenti in tecnologie di protezione dei dati, ma l'incidente ha messo in luce la vulnerabilità delle grandi aziende nel gestire enormi volumi di dati personali (Equifax, 2024)(FTC, 2024).

1.6.4 CoreLogic

Figura 5: Logo CoreLogic.

CoreLogic è un leader mondiale nella fornitura di dati e soluzioni analitiche per i settori immobiliare, finanziario e assicurativo. Fondata nel 1991, l'azienda ha sede a Irvine, California, e serve una clientela globale con un focus primario negli Stati Uniti, Canada, Australia e Nuova Zelanda. CoreLogic è nota per il suo vasto database che copre oltre il 99% delle proprietà residenziali negli Stati Uniti, rendendola una risorsa indispensabile per aziende del settore immobiliare e finanziario che necessitano di informazioni accurate su proprietà, transazioni e rischi catastali.

L'azienda genera oltre 1,9 miliardi di dollari di fatturato annuo e offre una vasta gamma di servizi di analisi, consulenza e gestione del rischio. Le sue soluzioni vengono utilizzate per migliorare la gestione dei portafogli immobiliari, ottimizzare la concessione di mutui e prestiti, e prevedere rischi associati al mercato immobiliare. CoreLogic raccoglie dati provenienti da fonti pubbliche e private, elaborandoli attraverso tecnologie avanzate come l'intelligenza artificiale e il machine learning per offrire previsioni accurate sui prezzi delle proprietà, sui

rischi di mercato e su potenziali frodi nel settore dei mutui. I dati raccolti comprendono informazioni su proprietà immobiliari, mutui, transazioni, rischi catastrofici e valutazioni finanziarie. Questi dati sono utilizzati da banche, compagnie assicurative, e agenzie governative per valutare la solvibilità dei mutuatari e ottimizzare le loro decisioni finanziarie.

Un esempio chiave della tecnologia di CoreLogic è la Discovery Platform, che consente ai professionisti di real estate e finanza di visualizzare e analizzare rapidamente i dati immobiliari attraverso una piattaforma integrata. Grazie alla precisione dei suoi dati, CoreLogic offre la possibilità di analizzare le tendenze di mercato e prevedere cambiamenti futuri con notevole accuratezza.

CoreLogic serve una vasta gamma di settori, tra cui:

- Real Estate: CoreLogic fornisce informazioni fondamentali a migliaia di agenti immobiliari, broker e società di gestione immobiliare, migliorando la trasparenza delle transazioni e delle valutazioni.
- Settore Finanziario: Le banche e le istituzioni finanziarie utilizzano i dati di CoreLogic per valutare il rischio di prestiti immobiliari e migliorare la gestione del rischio legato a mutui e linee di credito.
- Assicurazioni: CoreLogic aiuta le compagnie assicurative a stimare i rischi associati alle proprietà e a prevenire frodi attraverso dati dettagliati sui rischi catastali e sulle condizioni immobiliari.

Uno degli strumenti principali di CoreLogic è il suo Home Price Index (HPI), che fornisce valutazioni mensili sul prezzo delle case, utilizzato da istituzioni finanziarie e investitori per analizzare i movimenti del mercato immobiliare e prendere decisioni basate sui dati.

Gli enormi database contengono informazioni su oltre 99% delle proprietà residenziali negli Stati Uniti e su un'ampia gamma di proprietà in Canada, Australia e Nuova Zelanda. L'azienda raccoglie, archivia e analizza oltre 1 miliardo di proprietà a livello globale, contribuendo a creare un mercato immobiliare più trasparente e sicuro. CoreLogic vanta oltre 22.000 fonti di dati diverse, che alimentano i suoi strumenti di analisi per aiutare i clienti a prendere decisioni tempestive e accurate.

L’azienda ha affrontato critiche per l'eccessiva dipendenza da tecnologie di monitoraggio dei rischi immobiliari e dei mutui, che alcuni osservatori considerano vulnerabili a errori

algoritmici o al malfunzionamento in situazioni di mercato imprevedibili (Christl, 2017)(Corelogic, 2024).

1.6.5 Oracle

Figura 6: Logo Oracle.

Oracle Corporation è una delle aziende leader mondiali nel settore della tecnologia, specializzata nello sviluppo di software per la gestione dei database, sistemi cloud e soluzioni aziendali integrate. Fondata nel 1977 da Larry Ellison, Bob Miner e Ed Oates, Oracle è cresciuta fino a diventare uno dei principali fornitori globali di tecnologie per le imprese, con una particolare attenzione allo sviluppo di software di gestione dei dati, cloud computing, e infrastrutture IT avanzate.

Oracle offre una vasta gamma di prodotti e servizi, inclusi database, middleware, software aziendali (come ERP e CRM), nonché infrastrutture di cloud computing. La sua piattaforma di database, Oracle Database, è ampiamente utilizzata in tutto il mondo per gestire grandi quantità di dati e potenziare applicazioni significative per le aziende. Negli ultimi anni, Oracle ha investito significativamente nel cloud computing, espandendo le sue offerte di infrastruttura come servizio (IaaS), piattaforma come servizio (PaaS) e software come servizio (SaaS) attraverso Oracle Cloud.

Con sede centrale ad Austin, Texas, Oracle opera in oltre 175 paesi, servendo un'ampia gamma di clienti nei settori finanziario, sanitario, manifatturiero, delle telecomunicazioni, e governativo. L'azienda genera più di 40 miliardi di dollari di entrate annuali, e continua a crescere attraverso acquisizioni strategiche e innovazioni nel cloud computing e nella gestione dei dati.

Una delle divisioni di Oracle che opera direttamente nel campo del brokeraggio dei dati è la Oracle Data Cloud, che fornisce soluzioni di gestione dei dati utilizzate principalmente per

scopi di marketing digitale e pubblicità mirata. Questa divisione è stata potenziata dall'acquisizione di BlueKai nel 2014, un'importante piattaforma di gestione dei dati (DMP), specializzata nella raccolta e gestione di enormi set di dati sui consumatori per ottimizzare le campagne pubblicitarie digitali.

Oracle Data Cloud consente alle aziende di raccogliere dati provenienti da diverse fonti, inclusi dati di navigazione web, dati comportamentali, e dati transazionali. Grazie all'acquisizione di BlueKai, Oracle può combinare dati online e offline per creare profili dettagliati dei consumatori, questa capacità di raccogliere e unificare i dati provenienti da molteplici fonti fa di Oracle uno dei principali attori nel settore del brokeraggio dei dati.

Oracle Data Cloud raccoglie e vende dati relativi a:

- Dati demografici: età, sesso, reddito, stato civile.
- Dati comportamentali: cronologia di navigazione, acquisti online, interazioni con contenuti digitali.
- Dati geografici: localizzazione basata su IP e dati GPS.
- Dati transazionali: cronologia degli acquisti, partecipazione a programmi di fidelizzazione, e metodi di pagamento.

Oracle Data Cloud serve principalmente aziende che operano nel settore della pubblicità digitale, aiutandole a migliorare l'efficacia delle loro campagne di marketing. I clienti di Oracle includono inserzionisti digitali, piattaforme di e-commerce, e fornitori di contenuti che utilizzano i dati per migliorare la customer experience e aumentare il ritorno sugli investimenti pubblicitari. Le soluzioni di Oracle Data Cloud permettono di personalizzare le offerte pubblicitarie in base ai comportamenti e alle preferenze dei consumatori, migliorando significativamente l'efficacia delle campagne.

A livello globale, Oracle Data Cloud opera in diversi mercati, con una forte presenza in Nord America, Europa e Asia. Le sue soluzioni di gestione dei dati sono utilizzate da aziende leader nei settori del retail, delle telecomunicazioni, della finanza e dell'intrattenimento. Un esempio di successo è la collaborazione di Oracle con Procter & Gamble, che utilizza i dati raccolti per ottimizzare le sue strategie pubblicitarie a livello globale.

Oracle Data Cloud contribuisce in modo significativo alle entrate di Oracle nel settore dei big data e della pubblicità digitale. Tutto ciò ha permesso a Oracle di posizionarsi come uno dei leader del settore del brokeraggio dati, un mercato che continua a crescere rapidamente con

l'aumento dell'importanza dei dati comportamentali e delle tecnologie di machine learning per migliorare le strategie di marketing .

Nel 2020, Oracle è stata coinvolta in uno scandalo riguardante la vendita di dati di localizzazione raccolti tramite app mobili, sollevando ulteriori dubbi sull'uso dei dati per scopi pubblicitari (Ménard, 2023)(Oracle, 2024)(Dignan, 2014).

1.6.6 LexisNexis

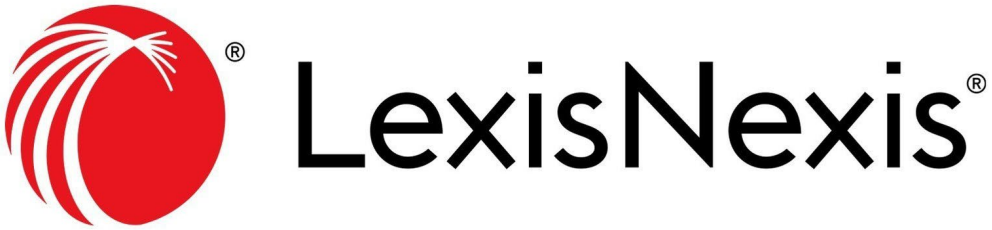

Figura 7: Logo LexisNexis.

LexisNexis è un'azienda globale specializzata nella fornitura di soluzioni di analisi dei dati, informazioni giuridiche e servizi di ricerca per un'ampia gamma di settori, tra cui legale, assicurativo, finanziario, sanitario e governativo. Parte del gruppo RELX, LexisNexis è stata fondata nel 1973 con l'obiettivo di creare una banca dati giuridica elettronica. Da allora, ha ampliato enormemente il proprio campo d'azione, diventando uno dei principali data broker a livello mondiale. Utilizza avanzate tecnologie di analisi dei dati, machine learning e intelligenza artificiale per raccogliere, organizzare e fornire informazioni a una vasta gamma di clienti, tra cui studi legali, aziende, istituzioni finanziarie, agenzie governative e università.

LexisNexis offre una varietà di servizi attraverso diverse industrie:

1. Settore Legale: Uno dei pilastri principali di LexisNexis è la sua ampia gamma di risorse legali poiché fornisce accesso a una vasta collezione di documenti legali, casi giuridici, statuti, regolamenti e notizie, diventando una risorsa imprescindibile per avvocati, giudici e professionisti del settore legale. Il suo database permette ai professionisti di condurre ricerche approfondite, accedere a informazioni storiche e ottenere informazioni su precedenti giuridici per supportare i loro casi.

2. Assicurazioni: Nel settore assicurativo, LexisNexis offre strumenti di valutazione del rischio e rilevazione delle frodi, raccogliendo dati da diverse fonti, tra cui registri pubblici e database proprietari, per aiutare le compagnie assicurative a valutare i profili di rischio dei potenziali assicurati. Questo approccio basato sui dati permette agli assicuratori di determinare le coperture e i tassi premio più appropriati, migliorando la gestione del rischio e riducendo le frodi.

3. Servizi Finanziari: LexisNexis fornisce strumenti essenziali per il settore finanziario, attraverso il suo servizio Risk Solutions, l'azienda aiuta le istituzioni finanziarie a valutare l'affidabilità creditizia di individui e aziende. Combinando dati di credito con altre informazioni, come registri pubblici e dati commerciali, consente alle banche e agli istituti di credito di prendere decisioni di concessione del credito più informate e di gestire il rischio in modo efficace.

4. Sanità: Nel settore sanitario, LexisNexis offre soluzioni per la verifica dell'identità e la prevenzione delle frodi sanitarie, grazie ai suoi database, l'azienda aiuta gli operatori sanitari a verificare con precisione l'identità dei pazienti e a prevenire furti di identità medica. Ciò contribuisce a garantire l'integrità dei registri dei pazienti e a mantenere la conformità con le normative sanitarie vigenti.

5. Governo e Forze dell'Ordine: Le agenzie governative e le forze dell'ordine utilizzano i dati di LexisNexis per supportare indagini, tracciare individui e raccogliere informazioni critiche per i casi legali. L'azienda fornisce strumenti di ricerca avanzati che facilitano il monitoraggio degli individui, la raccolta di dati personali e la verifica delle identità.

6. Media e Giornalismo: I giornalisti e i professionisti dei media si affidano a LexisNexis per una copertura informativa completa e approfondita; grazie alla sua vasta raccolta di articoli di notizie e pubblicazioni fornisce agli operatori del settore uno strumento per condurre ricerche, verificare fatti e ottenere insight su una vasta gamma di argomenti ed eventi.

I report di LexisNexis sono progettati per fornire un quadro completo di un individuo, integrando una varietà di dati da fonti diverse. Questi report includono:

1. Informazioni personali e identificative: Nomi, indirizzi e informazioni di contatto per identificare l'individuo.

2. Dati da registri pubblici: Comprendono interazioni con il sistema legale, come cause giudiziarie, sentenze e pignoramenti, fornendo una visione della storia legale dell'individuo.

3. Dati relativi al credito: Dettagli sulle transazioni finanziarie, incluse cronologie dei pagamenti, conti di credito e debiti in sospeso, aiutano le istituzioni finanziarie a valutare l'affidabilità creditizia di una persona.

4. Valutazione del rischio e scoring: I report di LexisNexis includono un punteggio di rischio, calcolato sulla base di algoritmi complessi che sintetizzano il rischio potenziale di un individuo in vari contesti, come assicurazioni e concessione del credito.

Questi dati vengono costantemente aggiornati per assicurare che i report siano sempre attuali e accurati. LexisNexis aggrega queste informazioni da diverse fonti, tra cui registri pubblici, database proprietari e agenzie di credito, creando così una narrazione completa e integrata della storia finanziaria, legale e personale di un individuo (LexisNexis, 2024)(Saliha, 2023).

1.7 Data Breach e Problematiche Legate alla Sicurezza

Le problematiche legate alla sicurezza e ai data breach sono tra le principali preoccupazioni quando si parla di data broker poiché queste aziende raccolgono e gestiscono enormi quantità di informazioni personali, spesso sensibili, e diventano obiettivi primari per attacchi informatici. Un data breach si verifica quando le informazioni protette vengono esposte o rubate da soggetti non autorizzati. Le violazioni dei dati, ossia l'accesso non autorizzato a informazioni sensibili, rappresentano una minaccia non solo per i consumatori, ma anche per le aziende e le istituzioni che acquistano e utilizzano tali dati, e attualmente, il ruolo dei data broker si colloca al centro di molte di queste vulnerabilità.

1.7.1 Rischi di Sicurezza nei Data Broker

Uno dei rischi principali nel settore dei data broker è l'archiviazione di grandi volumi di dati sensibili in singoli database centralizzati. Questi database contengono informazioni che spaziano dai dati finanziari ai dati di localizzazione, fino a dettagli sanitari e preferenze di acquisto. Ogni violazione di questi database può esporre i dati di milioni di individui, con conseguenze che vanno dal furto d'identità alla compromissione della privacy personale. Le informazioni ottenute attraverso un data breach possono essere utilizzate per frodi finanziarie, attacchi mirati, o anche per attività di sorveglianza da parte di terzi non autorizzati (Kosinski, 2023).

Gli attacchi informatici a data broker, come quelli a Experian e Equifax, mostrano quanto possano essere pericolose le falle nella sicurezza. Nel caso di Equifax, ad esempio, nel 2017 è stato esposto uno dei più grandi data breach della storia, compromettendo i dati di circa 147 milioni di persone. Gli hacker sono riusciti ad accedere a informazioni personali come numeri di previdenza sociale, date di nascita, indirizzi e numeri di patente di guida (FTC, 2024).

1.7.2 Conseguenze dei Data Breach

Le conseguenze di un data breach possono essere molteplici: a livello personale, gli individui i cui dati vengono violati rischiano il furto d'identità, accessi non autorizzati ai loro conti bancari o l'uso fraudolento delle loro informazioni per attività illecite, mentre sul fronte aziendale, queste violazioni possono danneggiare gravemente la reputazione delle società coinvolte. Un esempio significativo è il caso di Experian nel 2015, quando un attacco informatico compromise i dati di 15 milioni di utenti, minando la fiducia dei consumatori nell'azienda (FTC, 2019).

A livello legale, i data breach possono comportare ingenti sanzioni per le aziende. Dopo il breach di Equifax, l'azienda ha accettato di pagare fino a 700 milioni di dollari in risarcimenti e multe, comprese spese legali e sanzioni imposte dalla Federal Trade Commission (FTC). Il caso ha rappresentato una pietra miliare per le normative sulla sicurezza dei dati, dimostrando la gravità delle sanzioni imposte alle aziende che non proteggono adeguatamente le informazioni dei consumatori (Leonhardt, 2019).

1.7.3 Manipolazione e Sfruttamento dei Dati

Un altro problema significativo è l'uso improprio dei dati da parte di attori malevoli. Ad esempio, durante lo scandalo di Cambridge Analytica, i dati raccolti attraverso piattaforme social come Facebook sono stati utilizzati per influenzare l'opinione pubblica e manipolare le elezioni politiche negli Stati Uniti e nel Regno (Unito Della Piazza, 2021). Questo evento ha mostrato come i dati raccolti dai broker possano essere utilizzati non solo per fini commerciali, ma anche per scopi politici e manipolativi (Twetman e Bergmanis-Korats, 2020).

1.7.4 Il Mercato Nero dei Dati

Oltre al commercio legale di dati, esiste un vasto mercato nero dove i dati rubati vengono venduti per scopi criminali. La vendita di pacchetti di dati personali, noti come "fullz" (Steel, 2019), è una pratica diffusa nel dark web, dove informazioni complete su individui, come numeri di previdenza sociale, indirizzi e informazioni bancarie, possono essere acquistate a prezzi relativamente bassi. Questo rende i data breach ancora più preoccupanti, poiché i dati

rubati vengono rapidamente diffusi e utilizzati per crimini finanziari e frodi (Twetman e Bergmanis-Korats, 2020).

Capitolo 2: Modello di business dei Data Broker

2.1 Tipologia di dati raccolti

Fino a non molti anni fa, la raccolta di dati personali si limitava a informazioni piuttosto basilari e limitate come nome, indirizzo e alcuni dati demografici. Tuttavia, ad oggi, la situazione è drasticamente cambiata, aziende come Acxiom, hanno sviluppato la capacità di raccogliere fino a 11.000 punti dati per ciascun individuo (Kemp, 2023). Questi dati non si limitano più a semplici dettagli demografici, ma includono anche dati comportamentali, transazionali, di localizzazione e persino previsioni sui comportamenti futuri degli utenti. Questo incremento di dati collezionati è stato reso possibile dall'incremento di informazioni che gli utenti inseriscono online, che spaziano dalle interazioni sui social media alle recensioni dei prodotti, fino alle preferenze manifestate sui vari siti di e-commerce. L'enorme quantità di dati generata quotidianamente dai comportamenti online ha permesso di creare profili utente sempre più dettagliati, ulteriormente arricchiti dai nuovi metodi di tracciamento come i cookie, il fingerprinting del browser e il monitoraggio attraverso i dispositivi mobili. Queste tecniche avanzate consentono di monitorare con estrema precisione le attività online degli utenti, registrando ad esempio pagine visitate o azioni compiute sul web, accumulando così una vasta mole di dati comportamentali (Christl, 2017). Oggi, grazie a queste informazioni i data broker sono capaci di elaborare profili incredibilmente completi e dettagliati, includendo anche comportamenti e preferenze di ogni individuo. Le principali tipi di informazioni raccolti, elencati in seguito, (Vigderman e Turner, 2023) offrono una panoramica profonda che va dai dettagli personali ai modelli di consumo; tutto ciò permette ai data broker di offrire ai loro clienti strumenti potentissimi per il targeting e la personalizzazione, su una scala prima inimmaginabile.

- Dati Demografici: I dati demografici raccolgono informazioni basilari come età, sesso, indirizzo, etnia, reddito, livello di istruzione e stato civile. Queste informazioni sono fondamentali per le aziende perché permettono di comprendere la composizione del proprio mercato e di suddividere i consumatori in segmenti distinti e ben definiti.

- Dati Comportamentali: I dati comportamentali, più dettagliati rispetto ai dati demografici, comprendono le azioni e le interazioni degli utenti con siti web, app e social media. Informazioni di questo tipo sono di grande valore per le aziende che mirano a sviluppare profili di consumatori dettagliati e a prevedere i loro comportamenti futuri.

- Dati Geografici: I dati geografici, ottenuti tramite GPS o dall'analisi degli indirizzi IP, forniscono informazioni cruciali sulla posizione fisica degli utenti. Questi dati vengono impiegati non solo per il targeting pubblicitario basato sulla localizzazione dei consumatori ma anche per analisi di mercato e pianificazione strategica, come per esempio decidere dove aprire nuovi punti vendita scegliendo aree con una maggiore concentrazione di potenziali clienti.

- Dati di Interazione Sociale: Con l'avvento dei social media, le informazioni sulle interazioni sociali degli utenti sono diventate fondamentali per i data broker. Questi dati comprendono post condivisi, "mi piace", commenti, reti di contatti e attività in gruppi e forum. Offrono una visione dettagliata delle opinioni, delle preferenze e delle influenze sociali degli individui.

- Dati Transazionali: Questi dati sono fondamentali per comprendere in dettaglio le attività di acquisto e le transazioni finanziarie dei consumatori. Includono specifiche sui prodotti acquistati, i metodi di pagamento impiegati, la frequenza con cui vengono effettuati gli acquisti e il valore medio delle transazioni. L'analisi dei dati transazionali è cruciale per identificare e interpretare i pattern di spesa dei consumatori, permettendo alle aziende di affinare e personalizzare le loro strategie di marketing, ottimizzare le offerte, e migliorare l'efficacia delle campagne promozionali. Questi dati, quindi, giocano un ruolo chiave nel migliorare l'engagement del cliente e aumentare la fidelizzazione attraverso offerte mirate e promozioni calibrate sulle abitudini di spesa specifiche.

- Dati Sensibili: Questa categoria comprende informazioni estremamente delicate e personali quali orientamento sessuale, convinzioni religiose e politiche degli individui.

- Dati sulla Salute: Comprendono informazioni sulle condizioni di salute, prescrizioni mediche, e visite ospedaliere. Questi dati sono molto sensibili e sono spesso regolamentati da specifiche leggi sulla privacy.

- Dati Biometrici: Questi includono impronte digitali, riconoscimento del viso e altri dati biometrici utilizzati per l'identificazione personale.

- Dati sui Veicoli: questi dati includono dettagli sui veicoli posseduti, le modalità di guida, eventuali violazioni del codice della strada e le polizze assicurative associate. Queste informazioni sono essenziali per analizzare comportamenti di guida e gestire rischi e coperture assicurative.

- Dati sul Patrimonio Immobiliare: questi dati includono dettagli sugli immobili posseduti, le transazioni effettuate e le relative valutazioni delle proprietà, tali informazioni sono fondamentali per analizzare il mercato immobiliare e gestire le operazioni legate agli asset immobiliari.

- Dati sull'Educazione: Questi dati riguardano il percorso formativo degli individui, includendo le istituzioni educative frequentate, i titoli di studio ottenuti e le specializzazioni acquisite.

- Dati di Consumo di Media: questi comprendono informazioni relative ai programmi televisivi guardati, alle riviste lette e ai siti web che si visitano regolarmente. Tali dati sono fondamentali per analizzare le preferenze e le abitudini di consumo mediatico degli utenti.

- Dati di Utilizzo dei Servizi: questi includono informazioni relative all'uso di servizi quali elettricità, acqua, internet e altri servizi a pagamento, questi dati sono essenziali per analizzare le abitudini di consumo e gestire l'efficienza e la distribuzione delle risorse.

- Dati di Viaggio: questi racchiudono dettagli sui viaggi compiuti, inclusi i voli prenotati, gli alloggi utilizzati e le destinazioni visitate. Questi dati sono preziosi per analizzare le preferenze di viaggio e le abitudini dei consumatori nel settore turistico.

- Dati di Presenza a Eventi: Questa categoria comprende informazioni relative alla partecipazione a eventi, conferenze o altre manifestazioni pubbliche. Questi dati aiutano

a comprendere le preferenze e l'engagement degli individui in ambito culturale e professionale.

- Dati sugli Abbonamenti: Dettagli relativi agli abbonamenti a servizi vari, che includono sia piattaforme online, come lo streaming di film o musica, sia servizi offline, come l'iscrizione a palestre o club, informazioni che sono fondamentali per analizzare le tendenze di consumo e le preferenze dei servizi tra gli utenti.

2.2 Metodi di raccolta dati

I data broker impiegano numerosi metodi e si avvalgono di varie fonti per raccogliere i dati personali degli utenti, una delle principali è senza dubbio lo smartphone, che ogni giorno registra informazioni dettagliate sulla personalità e sulle attività quotidiane degli utenti. Molti individui possiedono account Google, Apple o Microsoft, e gran parte delle loro informazioni e delle loro attività vengono tracciate e registrate da questi account, creando una piattaforma che permette di identificare gli utenti in modo preciso. Ad esempio, i "like" su Facebook possono rivelare molti aspetti della vita delle persone, come l'etnia, le convinzioni religiose, l'uso di alcol, droghe e sigarette, con un livello di accuratezza variabile. Anche le ricerche effettuate su Internet possono fornire informazioni personali, come l'occupazione o il livello di istruzione di un individuo. In alcuni studi i ricercatori canadesi sono riusciti addirittura a rilevare lo stato emotivo delle persone analizzando il modo in cui digitano sulla tastiera dei loro dispositivi con buoni risultati di accuratezza come si può osservare nella Figura 8 (Christl, 2017).

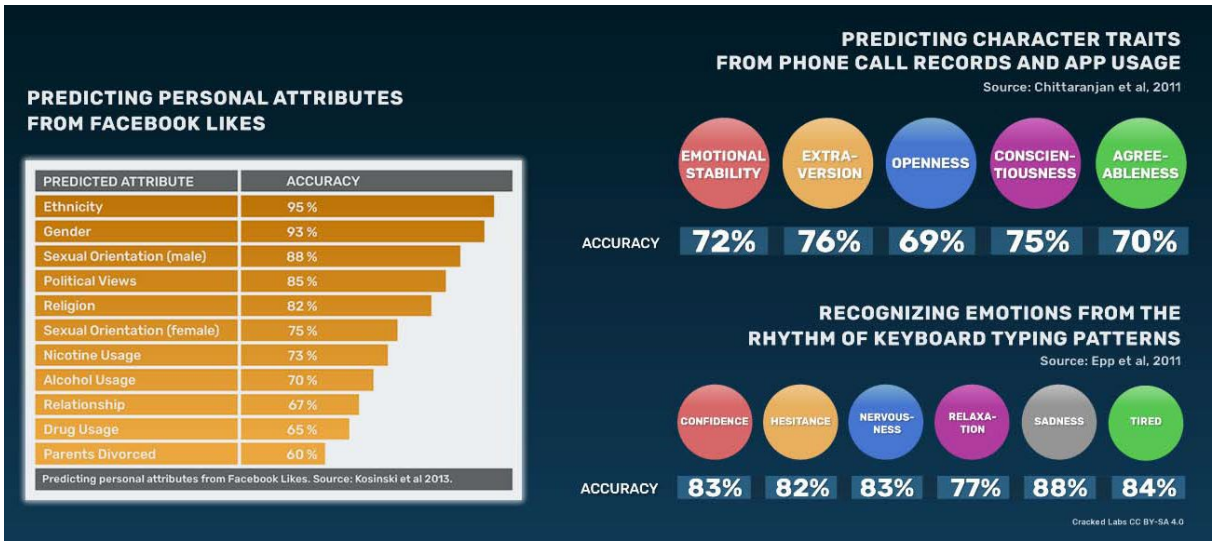

Figura 8: Analisi dell'accuratezza nella previsione di attributi personali e tratti caratteriali basati sui dati dei social media, registrazioni telefoniche e pattern di digitazione.

I Data Broker aggregano informazioni da una vasta gamma di fonti, sia online che offline, spesso senza rivelare dettagli specifici sulle origini dei dati. Ecco una panoramica generale, seppur non esaustiva, delle fonti da cui attingono (Vigderman e Turner, 2023) (Glasgow, 2018) (FTC, 2014):

- Aziende terze: uno dei metodi più comuni utilizzati dai data broker per raccogliere informazioni è la collaborazione con altre aziende. Possono lavorare con banche, agenzie di marketing o rivenditori, oppure acquistare i dati direttamente da loro. In alcuni casi, i broker si limitano a scambiare informazioni; ad esempio, grandi piattaforme online come Facebook, Twitter e Google collaborano con broker di dati per indirizzare la pubblicità e ottimizzarne l'efficacia.

- Registri pubblici: i data broker non si limitano al lavoro online; essi raccolgono anche i dati offline, infatti, riescono a raccogliere dettagli da registri immobiliari, documenti giudiziari, registri penali, registrazioni di veicoli, certificati di nascita, registri elettorali, dati del censimento, atti di matrimonio e divorzio e qualsiasi altro elenco o registro disponibile pubblicamente. Tutti i dati o le informazioni pubbliche sono aperti alla raccolta e all'uso; tuttavia, le leggi sulla privacy potrebbero impedire l'uso di questi dati in altri contesti, come il marketing e la propaganda.

- Attività online: i data broker monitorano le attività online degli utenti utilizzando diverse tecnologie di tracking, come i pixel di tracciamento e i cookie. La maggior parte

delle piattaforme online si avvale dei cookie per raccogliere e aggregare le attività degli utenti sul web, una pratica introdotta inizialmente dagli inserzionisti online. I cookie sono piccoli file di testo che i siti web installano sul dispositivo dell'utente durante la navigazione, vengono utilizzati per memorizzare diverse informazioni, come le preferenze dell'utente, il contenuto del carrello negli e-commerce o le credenziali di accesso. Oltre a migliorare l'esperienza dell'utente, i cookie vengono spesso condivisi con terze parti per tracciare il comportamento su vari siti web, consentendo agli inserzionisti di creare profili dettagliati e personalizzare gli annunci in base agli interessi specifici dell'utente. Sebbene gli utenti abbiano la possibilità di disattivare i cookie, molti finiscono per accettarli a causa del cosiddetto "sovraccarico di informazioni", un fenomeno che porta al "paradosso della privacy", in cui, nonostante la consapevolezza dei rischi per la propria privacy, si accettano comunque le condizioni per comodità o mancanza di alternative.

- Forniti dagli utenti. Questo avviene in diversi modi; spesso, quando si utilizza un'app, viene richiesto l'accesso a determinate informazioni personali (GPS, audio, contatti, ecc.) per poter usufruire del servizio ed in questo modo, i data broker possono raccogliere i dati senza che l'utente se ne accorga. Talvolta, la raccolta di dati avviene tramite una "strategia di incentivo", in cui le persone vengono persuase a fornire dati personali per partecipare a un concorso a premi (ad esempio, "iscriviti e vinci"). In alcuni casi, gli utenti condividono informazioni con piattaforme che, in realtà, operano anche come data broker nascosti. Alcuni broker possono talvolta raccogliere informazioni attraverso concorsi, questionari o sondaggi, sebbene questo metodo possa sembrare più trasparente, poiché i dati vengono inseriti direttamente dall'utente, sfrutta il fatto che molti non leggono attentamente i termini e le condizioni, ignorando il modo in cui le loro informazioni verranno successivamente utilizzate.

- Kit di Sviluppo Software (SDK). Gli SDK sono strumenti software che gli sviluppatori integrano nelle loro applicazioni per smartphone per aggiungere funzionalità o migliorare le prestazioni delle app. Questi kit svolgono anche un ruolo chiave nella raccolta di dati personali degli utenti e per questa ragione spesso i fornitori di questi SDK sono proprio i data broker che li offrono gratuitamente agli sviluppatori, con la promessa di migliorare la velocità e l'efficienza delle applicazioni. Tutto ciò avviene a discapito della privacy degli utenti, poiché una volta integrati nell'applicazione, gli SDK

possono raccogliere una vasta gamma di informazioni. Il vero problema di questa raccolta dati è che avviene in modo discreto, spesso all'insaputa sia degli utenti sia degli stessi sviluppatori, i quali integrano gli SDK senza conoscere appieno la quantità e la natura dei dati che verranno raccolti. Ciò accade perché i broker forniscono solo dettagli limitati sul funzionamento di questi strumenti in background. L'integrazione degli SDK nelle app, quindi, rappresenta una sorta di "cavallo di Troia" per i data broker, che ottengono accesso a una grande quantità di informazioni personali sotto l'apparenza di offrire un servizio gratuito e utile agli sviluppatori (Reviglio, 2022).

- Estrazione di dati tramite web: questa operazione si svolge principalmente attraverso due metodi: web scraping e data crawling. Con il web scraping, i data broker raccolgono automaticamente informazioni da specifici siti web utilizzando software noti come "scraper". Questi programmi simulano la navigazione umana, analizzando il codice HTML delle pagine web per individuare e raccogliere i dati di interesse, successivamente, gli scraper organizzano queste informazioni in un formato strutturato, come un foglio di calcolo o un database, consentendo così un'estrazione efficiente di grandi quantità di dati (CloudFlare, 2024). Il data crawling, invece, impiega tecniche simili ma con uno scopo più ampio: recuperare informazioni da qualsiasi tipo di fonte, non limitandosi solo al web. Queste tecniche sono legali, ampiamente diffuse e praticamente impossibili da evitare.

2.3 Aggregazione e Collegamento dei Dati Personali

I data broker combinano informazioni provenienti da molteplici fonti, sia online che offline, per costruire profili individuali estremamente dettagliati e ciò permette loro di ottenere una visione completa delle attività e delle caratteristiche di un utente, spesso all'insaputa dello stesso. I dati raccolti, che possono essere grezzi e presentare formati diversi, richiedono un accurato lavoro di riorganizzazione e pulizia prima di poter essere utilizzati in modo efficace. L’obiettivo finale è creare un profilo unico e completo, unendo informazioni spesso non correlate tra loro.

Un aspetto chiave di questa aggregazione è la capacità dei data broker di lavorare con dati strutturati, semi-strutturati e non strutturati.

- I dati strutturati sono quelli organizzati in formati predeterminati, come database e fogli di calcolo, con informazioni organizzate in righe e colonne (ad esempio, nomi, indirizzi e numeri di telefono).
- I dati semi-strutturati, invece, includono elementi organizzativi ma non seguono uno schema rigido, come i file JSON, XML o i metadati associati a documenti e immagini.
- Infine, i dati non strutturati rappresentano l'insieme più complesso, composto da contenuti come post sui social media, immagini, video, e-mail, articoli e trascrizioni di conversazioni, che non hanno un'organizzazione predefinita e richiedono tecnologie avanzate di analisi per estrarre informazioni utili.

Il processo di aggregazione inizia con la fase cruciale della pulizia dei dati, che mira a correggere errori, eliminare informazioni inaccurate o duplicate e garantire l'affidabilità e la precisione delle informazioni raccolte. Ad esempio, i dati sulla posizione possono variare notevolmente a seconda della fonte di provenienza. Per garantire l'utilità e l'accuratezza di queste informazioni, i data broker normalizzano i dati in un formato uniforme e coerente, indipendentemente dalla loro struttura originale.

Una volta puliti e standardizzati, i dati vengono sottoposti a un processo di collegamento dei profili, che consente di unire le informazioni provenienti da diverse fonti e creare un profilo completo e dettagliato di un individuo. Questo processo prevede l'identificazione di potenziali corrispondenze tra diversi profili e la valutazione della loro somiglianza in base a una serie di attributi chiave come nome, posizione geografica ed età. Attraverso questa complessa sequenza di riorganizzazione, pulizia e collegamento, i data broker riescono a trasformare una massa eterogenea di dati strutturati, semi-strutturati e non strutturati in profili individuali completi e di grande valore.

Nell'articolo "How data brokers endanger privacy" (Aïmeur et al., 2022) per illustrare concretamente la facilità con cui i data broker possono collegare dati provenienti da diverse fonti, gli autori del documento hanno progettato e implementato un sistema avanzato chiamato DROPLET. Questo sistema automatizzato ha la capacità di navigare attraverso una vasta gamma di siti di ricerca di persone, costruire in modo dinamico URL specifici per ciascun nome ricercato, e da questi estrarre una varietà di informazioni personali pubblicamente accessibili. Questo sistema funziona con una minima interazione umana e si basa su algoritmi sofisticati per standardizzare e collegare i dati raccolti da fonti multiple. L'algoritmo di funzionamento del sistema è il seguente:

1. Raccolta dei Dati (Data Collection)

- Costruzione degli URL**:** per raccogliere i profili, DROPLET visita una serie di siti web di ricerca di persone, costruendo URL specifici in base al nome fornito come input. Ad esempio, per cercare una persona chiamata "John Smith", costruisce l'URL appropriato per ciascun sito.
- Estrazione dei dati: accedendo ai siti, il sistema estrae informazioni personali disponibili gratuitamente come nome completo, età, indirizzo, numero di telefono, e-mail, posizioni passate, nomi dei parenti e altre informazioni sensibili. Ogni sito web può presentare i dati in modo diverso; quindi, il sistema è programmato per adattarsi a queste variazioni.
- Limitazioni: per aggirare le misure di sicurezza dei siti (come CAPTCHA), l'intervento umano è talvolta necessario per continuare la raccolta di dati.

2. Standardizzazione dei Dati (Data Cleaning)

- Prima di unire i profili, DROPLET pulisce e standardizza i dati raccolti per assicurarsi che i campi come luogo (città e stato) abbiano un formato uniforme. Ad esempio, rimuove le variazioni di formato nei campi come "Città, Stato" per assicurare una corretta corrispondenza tra profili da diversi siti.

3. Collegamento dei Profili (Data Linking)

- Scelta del database di riferimento: Prima di collegare i dati, DROPLET seleziona uno dei siti web come "database di riferimento", scegliendo il sito che ha il maggior numero di profili ricchi di informazioni.
- Blocco iniziale: Il sistema filtra i profili candidati in base a campi chiave come l'età e l'attuale luogo di residenza, eliminando quelli che non corrispondono.
- Calcolo della similarità: Per i profili che superano il filtro iniziale, il sistema calcola un punteggio di similarità confrontando campi come luoghi passati, relazioni (nomi dei parenti), e altre informazioni. I profili con punteggi di similarità più alti sono considerati corrispondenti.
- Selezione finale: DROPLET collega i profili da diversi siti al profilo nel database di riferimento che ha il punteggio di similarità più alto. In questo modo, può costruire un profilo più completo che unisce dati provenienti da fonti diverse.

4. Risultato Finale

- Il sistema fornisce come output profili “collegati” che combinano informazioni raccolte da diversi siti web di ricerca di persone. L'obiettivo principale è dimostrare quanto sia facile unire dati provenienti da diverse fonti, mettendo così in luce le potenziali minacce alla privacy.

Grazie al software DROPLET sviluppato dagli autori, che raccoglie dati personali da vari siti, li standardizza e poi li combina identificando le corrispondenze tra le diverse fonti, è stato possibile dimostrare quanto sia facile per i data broker aggregare una vasta quantità di informazioni. In questo modo, essi riescono a creare un profilo unico e completo per ogni individuo, combinando i dati raccolti in maniera dettagliata ed efficace.

2.4 Modelli di mercato, vendita e pricing dei dati

Le strategie adottate per commercializzare la vendita di dati usati dalle aziende spaziano dalla vendita diretta alla concessione di licenze e all'abbonamento tramite modelli Data-as-a-Service (Evelson et al., 2023). Ecco alcuni approcci specifici utilizzati dalle aziende:

- Vendita Diretta e API: Molte aziende possono offrire accesso ai loro dati tramite collegamento API, consentendo accesso sistematico o in tempo reale, il che facilita l'integrazione e l'utilizzo dei dati in applicazioni esterne. Alcune aziende vendono anche applicazioni che permettono agli utenti di visualizzare tendenze e insight derivati dai dati.

- Licenze e Modelli di Abbonamento: I dati vengono spesso commercializzati attraverso modelli di abbonamento, dove gli utenti pagano una tariffa regolare per accedere a set di dati aggiornati o a specifiche funzionalità analitiche. Questo modello è particolarmente prevalente nei servizi cloud e nei mercati digitali.

- Data as a Product (DaaP): Alcune aziende creano prodotti specifici a partire dai loro dati, combinando componenti di dati per offrire outcome di business monetizzabili. Questo può includere la combinazione di dati, algoritmi e API per fornire servizi specifici ai clienti.

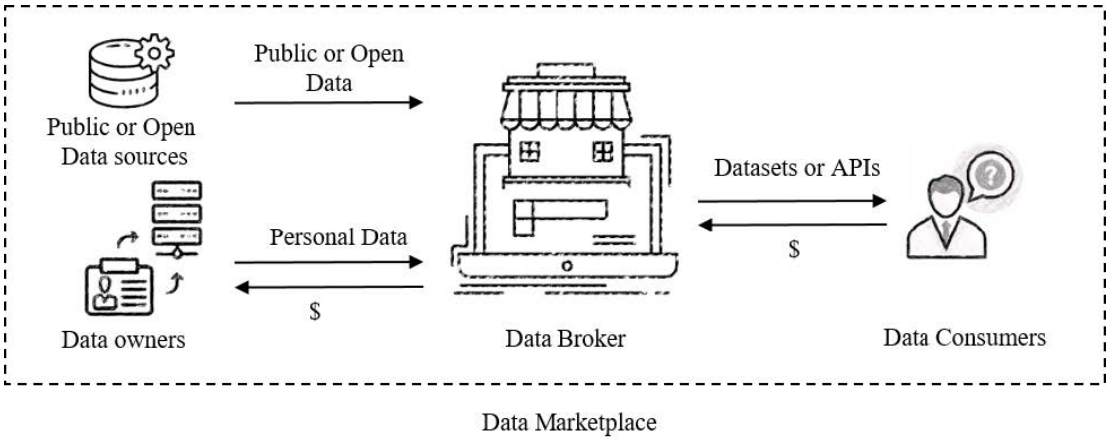

Figura 9: Schema di transazioni e scambi nel mercato dei dati.

Nel mercato, i data broker operano principalmente in due modi: come acquirenti e come venditori. Dal lato acquirente, i broker acquistano dati da aziende o individui che possiedono record su se stessi o sui propri utenti. Dal lato venditore, i broker offrono due tipi principali di prodotti: dataset e query. I dataset possono essere strutturati, come tabelle relazionali, o non strutturati, come immagini, audio, grafica e testo. La maggior parte dei dataset strutturati sono organizzati in tabelle dove ogni colonna rappresenta un attributo e ogni riga un record, associato a un oggetto specifico. Le query, invece, sono utilizzate per rispondere a domande specifiche poste dai consumatori di dati, le cui risposte possono essere delle viste, generalmente costituite da combinazioni di righe e colonne di un dataset (Beltrán et al., 2023). La struttura di mercato può essere descritta mediante l'uso di diversi modelli:

- One-side market: un modello di interazione di mercato in cui il databroker si interfaccia con acquirenti o con venditori, ma non contemporaneamente con entrambi.
    - Buy-side market: In questo mercato il data broker agisce da acquirente, come mostrato nella Figura, raccogliendo dati da più proprietari di dati. In questo modo serve organizzazioni e individui che desiderano monetizzare i propri dati.

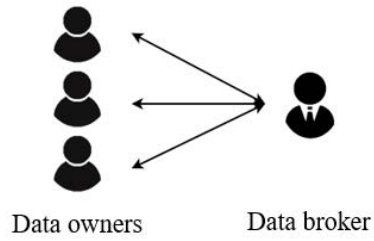

Figura 10: Buy-side market.

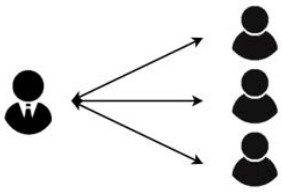

Figura 11: Sell-side market.

- Sell-side market: in questo scenario, il data broker opera come fornitore di dati (Figura). Qui, il data broker soddisfa le esigenze di organizzazioni e individui che, in qualità di "consumatori di dati", necessitano di queste informazioni per le loro attività.

- Two-sided market: questo modello di interazione di mercato coinvolge una piattaforma che serve sia il lato dell'offerta che quello della domanda. Un marketplace di dati funziona come un mercato bilaterale, dove il data broker, agendo da intermediario, mette a disposizione una piattaforma per lo scambio di dati tra chi possiede i dati e chi li utilizza, lucrando sulle transazioni effettuate.

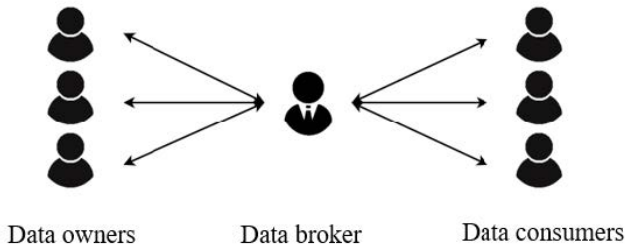

Figura 12: Centralised two-side market.

- Centralised two-side market: in questo modello di mercato, tutti gli scambi di dati tra proprietari e consumatori avvengono attraverso un data broker, come illustrato nella figura. Il data broker stabilisce sia i prezzi di acquisto per i fornitori di dati sia i prezzi di vendita per i consumatori, il suo deriva dalla differenza tra i ricavi ottenuti dalla vendita dei dati e i costi sostenuti per la loro raccolta. Poiché vengono venduti ai consumatori solo i risultati aggregati delle query, l'azienda garantisce la tutela della privacy personale dei fornitori di dati.

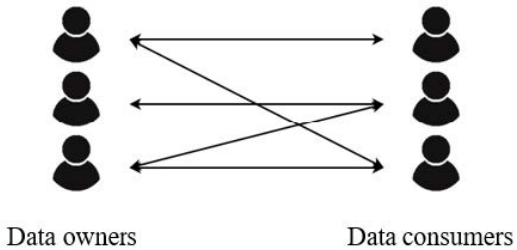

Figura 13: Decentralised two side-market.

- Decentralised two-side market: in questo tipo di mercato dei dati, come illustrato nella Figura, un data broker offre una piattaforma di trading per i proprietari e i consumatori di dati. Il broker non interviene direttamente nello scambio di dati, ma fornisce una piattaforma di scambio che è sotto il suo controllo. I consumatori e proprietari di dati interagiscono e

realizzano transazioni direttamente tra loro, purché siano membri della piattaforma di scambio. Datum, CitizenMe e DataWallet sono esempi di marketplace di dati che operano secondo questa struttura di mercato.

Le aziende devono affrontare diverse sfide quando decidono di commercializzare i loro dati, inclusa la definizione di un modello di consegna, la determinazione del valore dei dati per i clienti potenziali e l'adattamento del pricing in modo che rifletta il valore reale offerto. È cruciale adottare un approccio che metta il cliente al centro della strategia di commercializzazione dei dati per garantire il successo di tali iniziative.

2.4.1 Metodi di Pricing in simmetria di informazione

Nel caso di simmetria di informazioni possono essere usati diversi metodi di pricing nella vendita dei dati. Ciò è possibile se e sole se le preferenze dei consumatori e proprietari di dati sono note e in questo modo i data broker riescono a massimizzare il guadagno settando dei prezzi ottimali (Beltrán et al., 2023). In seguito, diversi metodi di pricing:

- Tariffa fissa: questo metodo di prezzo implica una quota di abbonamento solitamente mensile o annuale, che include un limite di utilizzo specifico. Nel caso di una tariffa fissa, il tempo è l'unico fattore che determinano il costo per il consumatore.

- Premium: il piano Premium è una variante speciale delle opzioni di abbonamento che prevede una quota mensile o annuale per un accesso illimitato. In questo modello, non vi sono limitazioni sul numero di dataset o query che un consumatore può eseguire.

- Pay-per-use (o prezzo lineare): Questo è un metodo di prezzo flessibile basato sul volume di utilizzo. Il costo p(n) è determinato linearmente in base alla quantità di utilizzo n, dove $p(n) = pu \times n$, e pu rappresenta il prezzo unitario. Confrontato con la tariffa fissa, il modello pay-per-use garantisce una maggiore flessibilità e risulta essere la scelta preferita per i consumatori occasionali di dati poiché permette di pagare basandosi sull'effettivo utilizzo, rendendolo ideale per chi non necessita di accesso costante ai dati.

- Two-part-tariff: questa tariffa combina elementi della tariffa fissa e del pay-per-use poiché i consumatori pagano prima una tariffa fissa per una quantità predefinita di

dataset o query. Se superano questa quota e desiderano più dati, possono acquistarli a un prezzo unitario, ovvero $p(n) = pf + pu \times n$, dove pf è la componente fissa e pu quella variabile. Questo modello si adatta bene a consumatori con esigenze di utilizzo variabili.

- Prezzo a livelli: Il prezzo a scalare definisce k prezzi unitari differenti basati sulle diverse quantità di un articolo. Questo sistema prevede k livelli di quantità tj e relativi prezzi pj. Le coppie di quantità e prezzi sono disposte come $(t1, p1)$; ...; (tk, pk). A ciascun livello, $p(m) = pjm$, dove $tj - 1 \leq m < tj$ per j da 2 a k, con $t1 \leq t2 \ldots \leq tk$ e $p1 \geq p2 \ldots \geq pk$. Man mano che la quantità aumenta, il prezzo unitario cala, riflettendo uno sconto per volumi che tende a incentivare le vendite. Nei sistemi di prezzo a scalare, occorre stabilire sia il prezzo pj che l'estensione dell'intervallo tra $tj - 1$ e tj.

- Versioning: il versioning consiste nel fornire più versioni di prodotti informativi a prezzi diversi. Le versioni possono variare in funzionalità, comodità e velocità di operazione e si rivolgono a diversi tipi di consumatori. Gli stessi dati possono essere venduti in più versioni con differenti qualità o quantità.

- Bundling (o Packaging): il bundling è un caso speciale di versioning poichè raggruppa più elementi insieme e li vende come un unico pacchetto. Il prezzo del pacchetto è solitamente inferiore alla somma dei prezzi degli articoli coinvolti acquistati singolarmente.

2.4.2 Metodi di Pricing in asimmetria di informazione

Nel caso di asimmetria informativa una delle parti detiene informazioni che le altre non possiedono, causando così un'allocazione inefficace delle risorse. Nel contesto di un mercato dei dati, dal lato della vendita, un consumatore di dati (ovvero un acquirente) conosce il valore atteso dei dati che intende acquistare, informazione che però non è accessibile al data broker e di conseguenza, il consumatore di dati può decidere di pagare un importo inferiore al valore effettivo per massimizzare il proprio beneficio. Per minimizzare le perdite economiche causate da questa asimmetria informativa, il data broker è incentivato a offrire certe condizioni vantaggiose per indurre i consumatori di dati a divulgare le informazioni che tengono riservate e quindi comprendere meglio il valore atteso dei dati per gli acquirenti. Sia i consumatori che i

proprietari di dati sono considerati agenti razionali che agiscono in modo strategico e ciò significa che, basandosi sulle regole di scambio stabilite, essi formulano ipotesi sulle strategie altrui e selezionano di conseguenza l'azione per essi più vantaggiosa. Per incentivare i proprietari di dati a rivelare il vero valore dei loro dati, il data broker deve garantire che questa trasparenza rappresenti per loro la strategia più vantaggiosa e che la partecipazione alle transazioni non comporti svantaggi (Beltrán et al., 2023). Le soluzioni più comuni per affrontare questo problema sono:

- Contratto: nei contratti, esiste una relazione lavorativa tra un agente e un principale, i cui interessi possono spesso essere divergenti. L'agente detiene maggiori informazioni rispetto al principale, il quale non può obbligarlo a divulgare tali informazioni. Il principale propone così all'agente una serie di contratti accuratamente strutturati tra cui scegliere, ognuno dei quali specifica un'azione che l'agente può compiere e la relativa compensazione. Analizzando le scelte e le azioni dell'agente, il principale può dedurre informazioni che altrimenti resterebbero occultate. Un esempio classico è il contratto assicurativo: la compagnia assicurativa, non conoscendo lo stato di salute del potenziale assicurato, propone vari contratti assicurativi, e, osservando la scelta del cliente, l'assicuratore è in grado di trarre conclusioni riguardo alla sua condizione di salute.

- Meccanismo incentivante: è un insieme di strategie analizzate nel contesto del design dei meccanismi, un approccio utilizzato per affrontare l'asimmetria informativa. Il design dei meccanismi può essenzialmente essere considerato come il processo inverso rispetto alla definizione degli equilibri di un gioco. Più specificamente, in un contesto di gioco in cui i partecipanti agiscono strategicamente per raggiungere certi obiettivi prefissati dagli equilibri del gioco stesso, il design dei meccanismi si propone di formulare le regole che influenzano questi risultati. Nella ricerca sul design dei meccanismi, due concetti principali di equilibrio sono particolarmente rilevanti: l'equilibrio di strategia dominante e l'equilibrio di Nash bayesiano. Con l'Equilibrio di Nash bayesiano, si assume che le preferenze individuali degli agenti siano informazioni private (nascoste agli altri agenti), ma che la distribuzione generale di queste preferenze nel mercato sia nota a tutti. Ogni agente, cerca di massimizzare la propria utilità attesa, basandosi sulle informazioni che ha riguardo alla distribuzione delle preferenze e ciò permette agli agenti di formulare strategie basate non solo sulle proprie informazioni ma anche sulle probabilità delle possibili preferenze degli altri. Mentre con l'equilibrio di strategia dominante, non si fanno ipotesi sulla conoscenza della distribuzione delle

preferenze degli altri agenti ma una strategia dominante è tale che, qualunque sia la scelta degli altri giocatori, la strategia in questione offre sempre il miglior risultato possibile per l'agente che la sceglie. In pratica, se un'agente ha una strategia dominante, la sceglierà perché massimizza la sua utilità a prescindere da cosa fanno gli altri.

- Aste: le aste sono un particolare tipo di meccanismo incentivante, utilizzato per determinare il prezzo e allocare beni o servizi ai partecipanti attraverso un processo competitivo di offerte. Esistono diversi tipi di aste, ciascuno con le proprie regole specifiche, ma tutte condividono l'obiettivo comune di scoprire il prezzo di mercato attraverso la concorrenza diretta tra i compratori. Grazie alla loro componente competitiva sono utili nel caso di asimmetria informativa come in questo caso, dove il valore dei dati non è chiaramente stabilito e può variare significativamente tra i diversi individui.

I data broker sono sempre alla ricerca di metodi innovativi per ottimizzare i loro modelli di pricing, in particolare quando si tratta di gestire e valutare dati non strutturati, che rappresentano una delle sfide più complesse in questo campo. Per affrontare queste problematiche, si avvalgono di tecniche avanzate di Machine Learning e Intelligenza Artificiale, che non solo facilitano la gestione di dati complessi ma permettono anche di sviluppare schemi di pricing dinamici e personalizzati. L'uso di questi algoritmi di apprendimento automatico aiuta i data broker a creare modelli di pricing che possono adattarsi in tempo reale alle variazioni del mercato e alle esigenze specifiche dei clienti. Inoltre, questa tecnologia offre la possibilità di simulare diversi scenari di pricing per determinare le strategie più vantaggiose prima di applicarle nel mercato reale. Questo approccio non solo aumenta l'efficienza e la competitività dei data broker ma migliora anche la soddisfazione del cliente, offrendo prezzi che riflettono meglio il valore dei dati in un contesto di mercato in continuo cambiamento.

2.5 Industria settori e dati

Il mercato dei data broker è estremamente ampio e variegato; queste aziende commercializzano dati a numerosi settori che li impiegano per ottimizzare le proprie operazioni, ridurre i rischi e potenziare offerte e servizi. Questa applicazione multidimensionale dei dati consente alle aziende di mantenere una posizione competitiva e di innovare continuamente. Uno studio

pubblicato nell'articolo *"How data brokers endanger privacy"* (Aïmeur et al., 2022) ha esaminato 75 data broker valutando i dati pubblici reperibili sulle pagine web dei data broker, esplorando vari attributi tra cui il settore di appartenenza e l'anno di fondazione. Mentre la maggioranza dei data broker esaminati offre servizi a imprese in cerca di soluzioni dati, 10 di essi, descritti come siti di ricerca di persone, si rivolgono agli individui in cerca di informazioni personali sui propri conoscenti, preferendo talvolta definire il loro business come "servizi al consumatore", dato che le loro attività sono orientate all'assistenza individuale. L'analisi ha rivelato che il settore più rappresentato è quello del Marketing & Advertising, che costituisce il 26% del totale, seguito da Information Technology & Services al 24% e Computer Software al 13%. Altri settori rilevanti includono Internet (13%) e i servizi di informazione (5%). Il grafico include anche una piccola percentuale di altri settori come i servizi ai consumatori e la ricerca di mercato, entrambi al 3%.

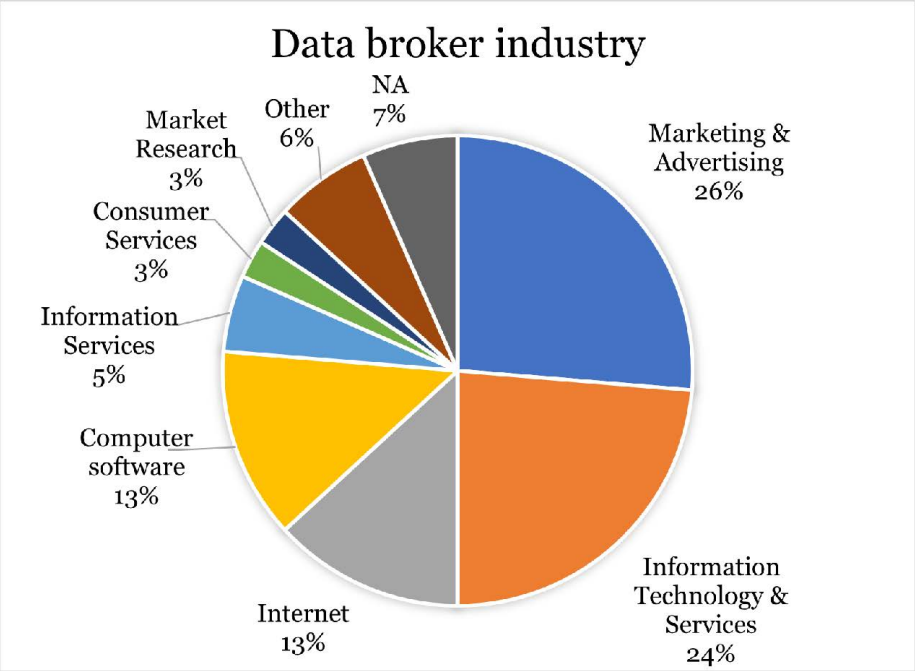

Figura 14: Settori di appartenenza dei data-broker analizzati.

Il grafico presentato in figura illustra l'evoluzione del settore dei data broker, mostrando il numero di nuove aziende emergenti nel mercato e raggruppandole per anno di fondazione. Le statistiche sono organizzate in intervalli decennali per offrire una visione chiara della crescita nel tempo. Dal grafico si evince un marcato incremento del numero di data broker a partire dagli anni '90, con circa un terzo dei broker analizzati che ha iniziato l'attività nell'ultimo decennio. Questo aumento riflette non solo il crescente interesse e l'importanza del settore dei

data broker ma anche la crescente attrattiva del mercato del brokeraggio dati, stimolata dall'avanzamento tecnologico e dall'espansione del Big Data.

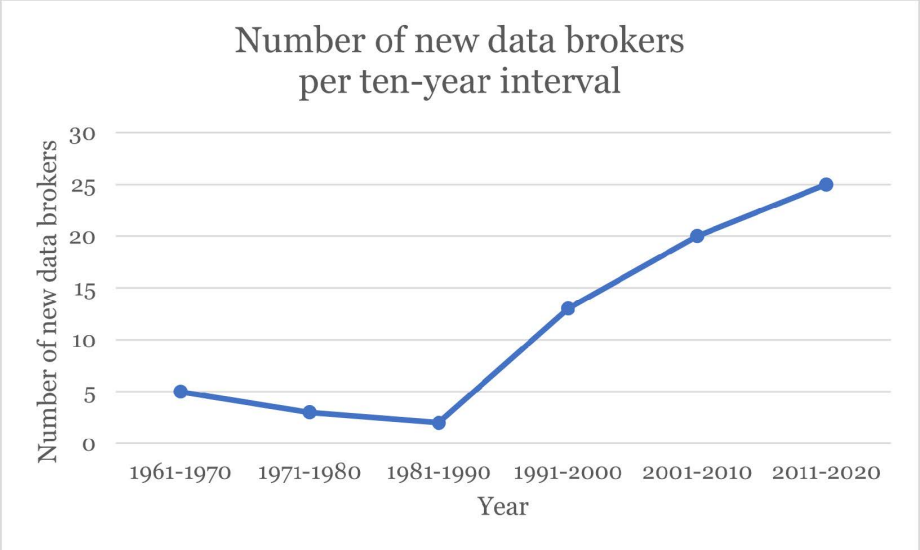

Figura 15: Numero di data-broker fondati per decennio.

Questa tendenza non solo dimostra la vitalità e il potenziale di investimento del settore dei data broker, ma riflette anche il loro ruolo crescente come intermediari essenziali in vari settori industriali (Sizelove, 2023). Di seguito, alcuni dei principali settori che si avvalgono dei servizi dei data broker:

- Aziende e organizzazioni: le aziende acquistano profili dettagliati dei clienti e dati di mercato per affinare le loro strategie di marketing. Questi dati permettono di personalizzare le offerte, segmentare il mercato in modo più efficace e migliorare il ROI delle campagne pubblicitarie attraverso una conoscenza approfondita delle preferenze e dei comportamenti dei consumatori.

- Ricerca medica: nel settore della ricerca medica, i dati sono fondamentali per condurre studi epidemiologici, sviluppare nuovi farmaci e monitorare gli esiti dei trattamenti. Questi dati aiutano a identificare la diffusione delle malattie, l'efficacia dei trattamenti esistenti e i potenziali rischi, accelerando significativamente il processo di innovazione e approvazione dei farmaci.

- Investigatori privati: gli investigatori privati utilizzano dati aggregati per condurre indagini su individui, eseguire verifiche di background o altre ricerche investigative; ciò può essere cruciale nella risoluzione di casi o per fornire sufficienti prove.

- Istituzioni finanziarie: le istituzioni finanziarie dipendono dai dati per valutare il rischio creditizio dei clienti, sviluppare nuovi prodotti finanziari e formare strategie di investimento.

- Agenzie assicurative: le assicurazioni utilizzano dati per determinare i premi assicurativi, valutare i rischi e sviluppare polizze personalizzate basate sui dati sulla salute, abitudini di guida, storia finanziaria e altro, per creare offerte che riflettano accuratamente il rischio individuale.

- Campagne politiche: le campagne politiche sfruttano i dati per individuare efficacemente gli elettori, pianificare strategie di campagna e comprendere le inclinazioni politiche delle diverse demografie permettendo così di ottimizzare le risorse e massimizzare l'impatto delle campagne.

- Agenzie federali: le agenzie federali utilizzano i dati per una varietà di scopi regolamentari, di monitoraggio e per comprendere l'efficacia delle politiche attuali, pianificando così interventi futuri basati su prove concrete.

- Compagnie farmaceutiche: queste compagnie utilizzano dati per la ricerca e sviluppo di nuovi farmaci, per condurre studi clinici e monitorare la sicurezza dei prodotti farmaceutici in commercio grazie all'accesso a vasti database di pazienti e risultati di trattamenti precedenti.

Ogni cliente sfrutta i dati per ottimizzare le proprie operazioni, ridurre i rischi e migliorare le proprie offerte o servizi. Questa crescente dipendenza da dati spiega il motivo per cui il fatturato dei data broker è notevolmente alto e in costante aumento: secondo l'analisi di Dataintelo (Dataintelo, 2024), il mercato globale dei data broker è stato stimato a 374,07 miliardi di dollari nel 2023 e si prevede che raggiungerà i 671,93 miliardi di dollari entro il 2032. Questo mercato è caratterizzato da un tasso di crescita annuo composto (CAGR) del 7,96% per il periodo 2024-2032. Questo incremento è attribuibile alla crescente richiesta di esperienze personalizzate per

i clienti a livello globale.

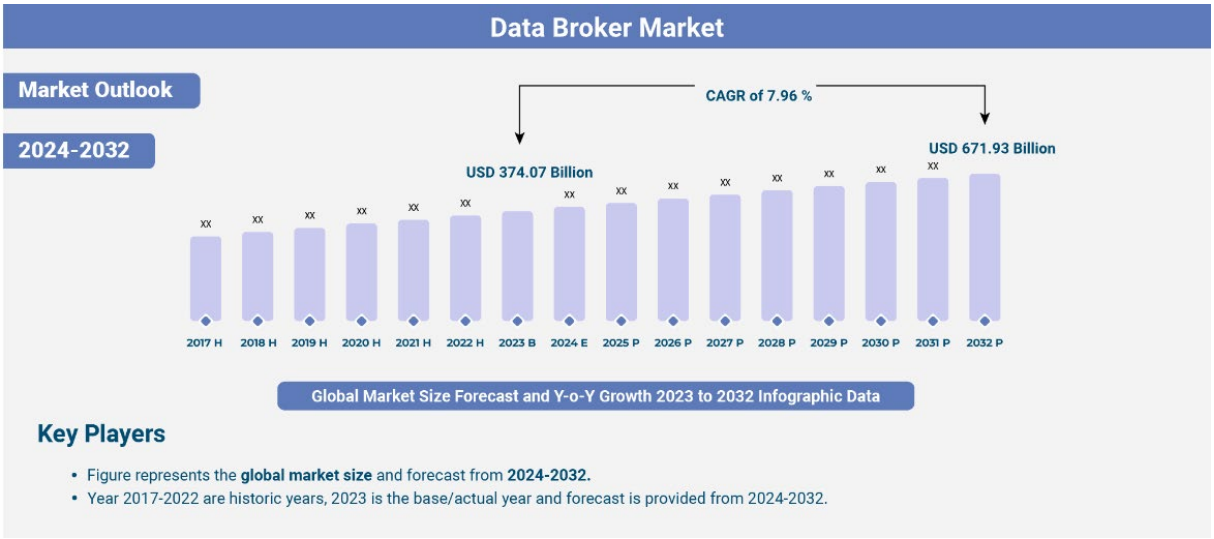

Figura 16: Volume economico del mercato dei data broker. Fonte: dataintelo.com

Capitolo 3: Regolamentazione dei Dati

3.1 Privacy dei dati

La privacy dei dati, nell'odierno panorama digitale, rappresenta un diritto fondamentale e una crescente preoccupazione per individui e organizzazioni in quest'era caratterizzata da una crescente raccolta e utilizzo dei dati. Si tratta del diritto di un individuo di avere il controllo sulle proprie informazioni personali e su come queste vengono raccolte, utilizzate e condivise da terzi. Come definito da IBM, "la privacy dei dati è la capacità di un individuo di controllare l'uso delle informazioni che lo identificano" (Kosinski e Forrest, 2023). Questo controllo si estende a diversi aspetti, tra cui:

- **Consenso**: gli individui devono poter dare il proprio consenso esplicito e informato alla raccolta e all'utilizzo dei propri dati, ciò implica che le organizzazioni devono essere trasparenti riguardo alle finalità del trattamento dei dati e fornire agli utenti la possibilità di scegliere quali informazioni condividere e per quali scopi.
- **Sicurezza**: le organizzazioni hanno la responsabilità di proteggere i dati personali da accessi non autorizzati, uso improprio o divulgazione. Ciò richiede l'implementazione di misure di sicurezza adeguate, che garantiscano la sicurezza dei dati già dalla raccolta.

- **Accuratezza**: le organizzazioni devono garantire la correttezza e l'accuratezza dei dati raccolti, poiché errori nei dataset possono provocare violazioni di privacy.
- **Trasparenza**: gli utenti hanno il diritto di sapere chi possiede i loro dati e come essi vengano utilizzati. Le organizzazioni devono comunicare chiaramente le informazioni raccolte e il loro scopo al momento della raccolta e informare gli utenti su eventuali cambiamenti nell'uso o condivisione dei dati con terze parti.
- **Cancellazione**: gli individui devono poter richiedere la cancellazione dei propri dati personali in determinate circostanze, come quando i dati non sono più necessari per le finalità per cui sono stati raccolti o quando l'individuo revoca il proprio consenso.

La privacy dei dati è strettamente collegata alla sicurezza, ma si focalizza su aspetti diversi. Mentre la sicurezza dei dati mira a proteggerli da minacce esterne, come attacchi informatici e violazioni, la privacy riguarda i diritti degli individui e il controllo che essi esercitano sulle proprie informazioni personali. Proteggere la privacy non è solo un diritto fondamentale che tutela la dignità e l'autonomia degli individui, permettendo loro di vivere senza il timore di essere monitorati o giudicati, ma è anche essenziale per prevenire abusi come furto di identità, discriminazione e manipolazione, poiché le informazioni personali, se non adeguatamente protette, possono essere utilizzate per scopi illeciti.

Per le organizzazioni, rispettare le normative in materia di dati e privacy non è solo un obbligo legale, ma un elemento determinante per costruire fiducia con gli utenti. La conformità a queste normative non solo evita sanzioni legali e danni reputazionali, ma rafforza la fiducia dei consumatori, incentivando la condivisione dei dati e la partecipazione all'ecosistema digitale stimolando così il processo digitale: quando le persone percepiscono che i loro dati sono trattati in modo sicuro e trasparente, sono più inclini a utilizzare servizi online.

L'attività dei data broker solleva preoccupazioni specifiche riguardo alla privacy e alla sicurezza dei dati. La loro attività, basata sulla raccolta, aggregazione e vendita di grandi quantità di dati personali, spesso sensibili, richiede una maggiore attenzione alla protezione dei diritti degli individui. L'opacità dei processi di raccolta e profilazione, insieme ai rischi di discriminazione e manipolazione, evidenzia l'importanza di normative solide sulla privacy per garantire la legittimità e l'accettazione delle loro operazioni. Solo una gestione responsabile dei dati può conciliare l'innovazione tecnologica con la tutela dei diritti fondamentali.

Infine, il rispetto delle normative sulla privacy non solo aiuta a evitare sanzioni e danni alla reputazione, ma rappresenta un vantaggio competitivo, perché in un contesto in cui la

consapevolezza della privacy è in crescita e gli scandali relativi alla violazione dei dati sono frequenti, essere percepiti come rispettosi della privacy può differenziare significativamente un data broker dalla concorrenza. Adeguarsi a regolamenti complessi e variabili richiede investimenti significativi in tecnologie di sicurezza e personale specializzato, influenzando direttamente i costi operativi e le strategie aziendali, rendendo la gestione dei dati un elemento chiave per il posizionamento competitivo sul mercato.

3.2 Information Fatigue e Privacy Paradox

Nell'ambiente moderno, l'ininterrotto flusso di informazioni digitali e l'uso prolungato di dispositivi tecnologici stanno influenzando negativamente la salute mentale e fisica delle persone. Il concetto di Information Fatigue Syndrome (IFS), delineato dallo psicologo britannico David Lewis, descrive un tipo di esaurimento risultante dall'essere costretti a processare un eccesso di informazioni in brevi lassi di tempo, generando sintomi come apatia, indifferenza e stanchezza mentale. Questi problemi sono aggravati dalla necessità di rimanere continuamente aggiornati, una caratteristica dominante del nostro saturato ambiente mediatico.

I materiali non filtrati attraverso i canali editoriali tradizionali, come blog, newsletter e rapporti non ufficiali, detti "grey literature", giocano un ruolo significativo nell'incrementare l'IFS. Questo tipo di contenuto, spesso privo di rigorosi controlli di qualità, aggiunge un ulteriore livello di complessità e sovraccarico informativo, rendendo ancora più arduo per gli individui distinguere tra dati affidabili e irrilevanti. Tale sovraccarico informativo può sfociare in quello che è stato descritto come il "paradosso della privacy", un fenomeno contraddittorio dove gli utenti, nonostante esprimano preoccupazioni significative per la sicurezza dei propri dati personali, finiscono per adottare comportamenti che compromettono la loro privacy stessa (Savić, 2023).

Nel dettaglio questo fenomeno viene chiamato “privacy fatigue” (Chen et al., 2022) una situazione in cui gli utenti mostrano un atteggiamento cinico verso la protezione dei dati, riducendo la motivazione a impegnarsi in comportamenti attivi di salvaguardia della privacy nonostante essi siano consapevoli dei rischi. Questo affaticamento, derivante dalla necessità di prendere costantemente decisioni su come e quando proteggere i propri dati personali, attenua l'effetto delle preoccupazioni sulla privacy sull'intenzione di adottare misure protettive e allenta gli utenti da comportamenti di difesa verso i rischi di sicurezza. Un esempio di questa situazione è quando gli individui, stanchi del bombardamento continuo di richieste di consenso e

aggiornamenti sulla privacy, scelgono di cliccare "accetto" senza leggere i dettagli dei termini di servizio. Questa pratica comune sottolinea le difficoltà degli utenti nel mantenere un controllo attivo e informato sui propri dati personali, in un mondo digitale caratterizzato da politiche e procedure spesso estese e complesse.

La letteratura accademica ha approfondito le basi psicologiche e comportamentali del paradosso della privacy, adottando spesso il "calcolo della privacy" (Gerber et al., 2018) come modello esplicativo dominante. Secondo questo approccio, si ritiene che gli utenti effettuino una valutazione istintiva tra i benefici immediati ottenibili dalla condivisione dei dati e i rischi a lungo termine, come il furto d'identità o la perdita di controllo sulle proprie informazioni. Questo processo decisionale, tuttavia, è spesso soggetto a bias cognitivi che possono alterare la percezione dei rischi e dei benefici, rendendo più complessa per gli utenti la capacità di fare scelte che rispecchino con precisione le loro preoccupazioni in termini di privacy. Le variazioni nei comportamenti legati alla privacy, a seconda dei contesti sociali e culturali, dimostrano l'importante impatto che le norme sociali e le pressioni dei pari possono avere sulle preoccupazioni e sulle azioni relative alla privacy. In certi contesti, il bisogno di conformarsi socialmente può indurre gli individui a rivelare più informazioni personali di quanto normalmente farebbero.

Per contrastare questi fenomeni, è efficace adottare strategie come la selezione attenta delle informazioni e il consumo di contenuti di alta qualità che arricchiscano piuttosto che sopraffare. È altrettanto essenziale inserire regolari pause dall'assorbimento di nuove informazioni per permettere alla mente di rigenerarsi e minimizzare il rischio di esaurimento. Queste pratiche suggeriscono un cambiamento nell'interazione con le tecnologie e nella gestione del flusso di informazioni, favorendo un approccio più consapevole e controllato, in questo modo si incrementa l'efficienza nel processare le informazioni in modo sostenibile, contribuendo alla tutela della privacy.

3.3 Data act

Il Regolamento (UE) 2023/2854 chiamato più comunemente Data Act è un regolamento europeo che disciplina l'accesso e dell'uso dei dati generati da prodotti e servizi connessi, come smartphone, smart TV e macchinari industriali. Con l'intento di creare un mercato dei dati più equo e competitivo, questa legge cerca di stimolare l'innovazione e la crescita economica attraverso una distribuzione equa del valore dei dati, definendo chiaramente chi può utilizzare

i dati e in quali condizioni. Ciò assicura un accesso equo a tutti gli attori del mercato, dalle grandi aziende alle piccole e medie imprese fino ai consumatori, smantellando così di fatto il monopolio che hanno le big tech sui dati.

Il Data Act promuove lo sviluppo di nuovi servizi e prodotti basati sui dati, soprattutto nell'ambito dell'Internet of Things (IoT), e mira a creare un mercato unico per i dati all'interno dell'Unione Europea, facilitando la libera circolazione dei dati, rimuovendo gli ostacoli e promuovendo l'interoperabilità tra diverse piattaforme e servizi. Il regolamento garantisce la protezione dei dati personali e previene gli abusi contrattuali, stabilendo norme chiare per la condivisione dei dati tra le imprese e consentendo alla Commissione di sviluppare clausole contrattuali standardizzate.

Gli utenti, sia consumatori che imprese, hanno il diritto di accedere ai dati che generano e di condividerli con terze parti dall'altro lato i produttori di dispositivi connessi sono tenuti a rendere i dati facilmente accessibili, e gli enti pubblici possono accedere ai dati detenuti dalle imprese in situazioni di interesse pubblico, come le emergenze. Il regolamento facilita anche la portabilità dei dati e promuove l'interoperabilità, migliorando così la competitività e l'efficienza delle imprese attraverso l'analisi dei dati.

Il Data Act, entrato in vigore l'11 gennaio 2024 con applicazione a partire da settembre 2025, rappresenta un passo fondamentale per creare un'economia dei dati europea più equa, competitiva e innovativa, a beneficio di tutti gli attori coinvolti. Con questo quadro regolamentare, l'Unione Europea mira a stimolare l'innovazione e la crescita economica, garantendo al contempo che la digitalizzazione proceda in modo equo e sostenibile (Commissione Europea, 2024).

3.4 GDPR

Il Regolamento (UE) 2016/679 più comunemente chiamato Regolamento Generale sulla Protezione dei Dati (GDPR) è una legge europea che regola il trattamento dei dati personali all'interno dell'Unione Europea. È entrato in vigore nel maggio 2018, sostituendo la precedente direttiva sulla protezione dei dati (direttiva 95/46/CE) e rappresenta una pietra miliare nella legislazione sulla privacy, introducendo un quadro normativo completo e armonizzato per la protezione dei dati personali di tutti i cittadini dell'UE. Insieme al più recente Data Act, costituisce il nucleo normativo fondamentale che regola e delinea in modo stringente le attività dei Data Broker all'interno dell'Unione Europea.

Il GDPR si propone di raggiungere diversi obiettivi fondamentali:

- Armonizzare le leggi sulla protezione dei dati in tutta l'UE: prima dell'introduzione del GDPR, ogni Stato membro dell'UE aveva le proprie leggi sulla protezione dei dati, il che creava frammentazione e incertezza. Il GDPR ha creato un quadro normativo unico e coerente per tutta l'Unione, semplificando la conformità per le aziende che operano in più paesi e garantendo una protezione uniforme per tutti i cittadini dell'UE.
- Rafforzare la protezione dei dati personali dei cittadini europei: il GDPR rafforza i diritti degli individui in relazione ai loro dati personali, garantendo loro maggiore controllo su come le loro informazioni vengono raccolte, utilizzate e condivise.
- Aumentare la responsabilità e la trasparenza dei titolari del trattamento dei dati: il GDPR pone una maggiore enfasi sulla responsabilizzazione delle organizzazioni che trattano dati personali. Le aziende devono adottare misure proattive per proteggere i dati personali e dimostrare la loro conformità al regolamento.
- Adattare la legislazione all'era digitale: il GDPR tiene conto delle nuove sfide poste dalle tecnologie digitali, come l'intelligenza artificiale, il cloud computing e l'Internet delle cose (IoT), e fornisce un quadro normativo per il trattamento dei dati personali in questi contesti.
- Promuovere la fiducia nell'economia digitale: il GDPR mira a creare un ambiente di fiducia in cui gli individui si sentano sicuri di condividere i loro dati online, favorendo lo sviluppo dell'economia digitale.

L'articolo 2 del GDPR specifica le situazioni in cui il regolamento è applicabile in termini di trattamento dei dati personali. Esso si applica al trattamento dei dati personali che è parzialmente o totalmente automatizzato, e anche al trattamento non automatizzato di dati personali che sono parte di un archivio o destinati a farne parte. Si escludono però le seguenti modalità di trattamento di dati personali:

- Quando eseguito da una persona fisica nell'ambito di attività personali o domestiche.
- Se relativo alla sicurezza nazionale o alla difesa.
- Quando eseguito da autorità competenti per fini di prevenzione, indagine, accertamento o persecuzione di reati o esecuzione di sanzioni penali.

L'articolo 3 del GDPR estende l'applicazione del regolamento non solo alle attività che si svolgono all'interno dell'UE, ma anche a soggetti non stabiliti nell'UE quando le loro attività riguardano l'offerta di beni o servizi a persone nell'UE, o il monitoraggio del loro

comportamento, purché tale comportamento avvenga all'interno dell'UE. Questo significa che, indipendentemente dalla localizzazione del trattamento, se le attività sono dirette verso individui nell'UE, il GDPR sarà applicabile. Il GDPR si applica anche al trattamento di dati personali effettuato da istituzioni, organi, e agenzie dell'Unione, indipendentemente dal luogo di trattamento.

Il GDPR è applicato dalle autorità di controllo nazionali in ciascuno Stato membro dell'UE. Queste autorità sono responsabili di sorvegliare l'applicazione del regolamento, fornire consulenza alle organizzazioni e agli individui e gestire i reclami degli interessati. In Italia, l'autorità di controllo è il Garante per la protezione dei dati personali ed ha potere di:

- condurre indagini e ispezioni presso le organizzazioni che trattano dati personali;
- emettere avvertimenti e sanzioni amministrative pecuniarie alle organizzazioni che violano il GDPR;
- ordinare la sospensione o il divieto del trattamento dei dati;
- fornire consulenza e assistenza alle organizzazioni e agli individui in materia di protezione dei dati.

Gli individui che ritengono che i loro diritti siano stati violati possono presentare un reclamo al Garante o per vie legali.

Il GDPR impone alle aziende che trattano dati personali una serie di obblighi rigorosi per garantire la protezione e la corretta gestione di tali informazioni. Tra questi obblighi figura la necessità di basare il trattamento dei dati su una fondazione legale chiara, come il consenso dell'interessato o l'adempimento di un contratto. Le aziende sono anche tenute a operare con trasparenza, fornendo agli interessati informazioni chiare su come i loro dati vengono gestiti. Inoltre, le aziende devono limitarsi a raccogliere dati strettamente necessari per lo scopo specificato, assicurandosi altresì che questi siano accurati e mantenuti solo per il tempo necessario. È fondamentale che vengano adottate misure di sicurezza per proteggere i dati da accessi non autorizzati o trattamenti illeciti e dimostrare attivamente la conformità a queste regole.

Per le aziende che rientrano in certe categorie, come quelle che effettuano trattamenti su larga scala o sono autorità pubbliche, è richiesta la nomina di un Responsabile della Protezione dei Dati (DPO). Queste aziende devono anche condurre Valutazioni d'Impatto sulla Protezione dei Dati (DPIA) per trattamenti che presentano rischi elevati. Ogni violazione dei dati che possa

compromettere i diritti e le libertà degli individui deve essere notificata all'autorità di controllo e, in certi casi, agli interessati stessi.

A livello di obblighi specifici del Titolare del Trattamento, il GDPR articola nei suoi articoli vari principi e requisiti:

- L'Articolo 5 stabilisce principi chiave come la liceità, la trasparenza, la minimizzazione dei dati, l'accuratezza, la limitazione della conservazione, l'integrità e la riservatezza.
- L'Articolo 6 delinea le condizioni sotto le quali il trattamento dei dati è considerato legale.
- L'Articolo 12 richiede che le informazioni sul trattamento siano fornite in modo trasparente e comprensibile.
- L'Articolo 30 obbliga i titolari e i responsabili del trattamento a mantenere un registro delle attività di trattamento.
- Gli Articoli 33 e 34 delineano le procedure per la notifica delle violazioni dei dati personali sia alle autorità di controllo che agli interessati.

Questi articoli si combinano per formare un quadro normativo che richiede alle aziende di trattare i dati personali con la massima cura, assicurando al contempo che gli interessati siano informati e in grado di esercitare i loro diritti.

Il GDPR nell’articolo 9 tratta le categorie di dati personali considerate sensibili con norme particolarmente stringenti, riconoscendo la loro natura delicata. Queste categorie particolari includono informazioni su origine razziale o etnica, opinioni politiche, convinzioni religiose o filosofiche, appartenenza sindacale, nonché dati genetici, biometrici per identificazione univoca, dati sulla salute, e dettagli riguardanti la vita sessuale o l'orientamento sessuale di una persona. Il trattamento di tali dati è generalmente proibito, a meno che non siano soddisfatte condizioni specifiche come il consenso esplicito e informato dell'interessato per scopi definiti. Altre eccezioni includono situazioni in cui il trattamento è necessario per obblighi legali specifici in materia di lavoro, per la protezione degli interessi vitali dell'interessato, durante attività legittime di associazioni non profit, o per la gestione di sistemi sanitari o sociali.

Norme specifiche possono applicarsi nel contesto lavorativo, dove gli Stati membri possono introdurre regolamenti più dettagliati per il trattamento dei dati personali dei dipendenti, inclusi quelli sensibili (Articolo 88).

Per quanto riguarda il processo decisionale automatizzato, inclusa la profilazione che utilizza categorie di dati sensibili, il GDPR impone restrizioni severe, consentendo tali attività solo sotto condizioni strettamente controllate, come il necessario completamento di un contratto, la protezione garantita da leggi appropriate, o sulla base di un consenso chiaro e esplicito (Articolo 22).

Il GDPR riconosce e protegge vari diritti degli interessati relativi al trattamento dei loro dati personali. Tra questi, il diritto di accesso (Articolo 15) permette agli interessati di richiedere e ottenere dal titolare del trattamento la conferma dell'esistenza di un trattamento dei propri dati personali, con l'accesso ai dati e informazioni aggiuntive riguardanti gli scopi del trattamento e i destinatari dei dati.

Gli interessati hanno anche il diritto di rettifica (Articolo 16), che consente loro di correggere dati inaccurati senza ingiustificato ritardo. Strettamente correlato è il diritto all'oblio (Articolo 17), attraverso il quale possono richiedere la cancellazione dei loro dati personali sotto determinate condizioni.

Inoltre, il diritto di limitazione del trattamento (Articolo 18) offre agli interessati la possibilità di limitare il trattamento dei propri dati in specifiche circostanze. Il diritto alla portabilità dei dati (Articolo 20) consente agli interessati di ricevere i propri dati in un formato strutturato e di trasferirli ad altro titolare del trattamento.

Infine, il diritto di opposizione (Articolo 21) permette agli interessati di opporsi al trattamento dei loro dati per motivi legati alla loro particolare situazione, fornendo un controllo significativo sull'uso dei propri dati personali. Questi diritti assicurano che gli individui non solo siano informati del trattamento dei loro dati, ma abbiano anche mezzi efficaci per controllare tale trattamento.

Il non rispetto del GDPR può portare a severe sanzioni contenute nell’articolo 83 e nell’articolo 84 che stabiliscono le condizioni e i criteri per l'imposizione di sanzioni amministrative pecuniarie, che possono essere molto severe. La gravità delle sanzioni dipende dalla natura, gravità e durata della violazione, nonché dall'intenzionalità, dalle misure adottate per mitigare il danno, dal numero di persone interessate, e da eventuali violazioni precedenti, tra altri fattori. Le multe possono raggiungere fino a 20 milioni di euro o, nel caso di

un'impresa, fino al 4% del fatturato annuo globale totale dell'anno precedente, a seconda di quale sia maggiore. Tutto ciò ha spinto le aziende a investire significativamente in compliance, migliorando così le pratiche di gestione dei dati e la trasparenza poiché la conformità al regolamento richiede un impegno continuo e un'attenta valutazione dei rischi per la privacy, ma a lungo andare può portare a benefici significativi in termini di reputazione, competitività e sostenibilità aziendale.

Il GDPR ha un impatto significativo sulle attività dei data broker, dato che la loro attività si basa essenzialmente sulla raccolta di grandi quantità di dati personali; lo fa imponendo loro obblighi specifici e rafforzando i diritti degli interessati.

Alcune delle principali implicazioni del GDPR per i data broker includono:

- Base giuridica per il trattamento: i data broker devono dimostrare di avere una base giuridica valida per il trattamento dei dati personali, come il consenso esplicito dell'interessato o la necessità di adempiere a un obbligo contrattuale o legale. L'articolo 6 del GDPR elenca le diverse basi giuridiche per il trattamento dei dati.
- Trasparenza: i data broker devono fornire informazioni chiare e concise agli interessati sulle modalità di trattamento dei dati, inclusi la finalità del trattamento, le categorie di dati trattati, i destinatari dei dati e i diritti degli interessati. L'articolo 12 del GDPR stabilisce gli obblighi di trasparenza per i titolari del trattamento.
- Diritti degli interessati: i data broker devono garantire agli interessati l'esercizio dei loro diritti, come il diritto di accesso ai propri dati, il diritto di rettifica dei dati inesatti, il diritto alla cancellazione dei dati ("diritto all'oblio") e il diritto alla portabilità dei dati. Il Capo III del GDPR elenca i diritti degli interessati.
- Sicurezza dei dati: i data broker devono adottare misure tecniche e organizzative adeguate per proteggere i dati personali da accessi non autorizzati o trattamenti illeciti, come la cifratura dei dati, la pseudonimizzazione e l'autenticazione a più fattori. L'articolo 32 del GDPR stabilisce gli obblighi di sicurezza per i titolari del trattamento.
- Data Protection Impact Assessment (DPIA): i data broker devono effettuare una DPIA (disciplinata nell’articolo 35 del GDPR) per i trattamenti che possono presentare un rischio elevato per i diritti e le libertà degli interessati, come la profilazione o il

trattamento di dati sensibili. La DPIA è un processo che valuta l'impatto del trattamento sulla privacy degli individui e aiuta a identificare e mitigare i rischi.

3.5 Regolamentazioni in USA

Negli Stati Uniti non esiste una legge federale generale che regoli la raccolta e la vendita di dati da parte di entità commerciali, inclusi i data broker, né tantomeno un equivalente diretto del GDPR europeo, ovvero una normativa comprensiva per la protezione dei dati personali. Esistono invece diverse leggi federali che proteggono tipi specifici di informazioni, come la Health Insurance Portability and Accountability Act (HIPAA) per i dati sanitari e la Family Educational Rights and Privacy Act (FERPA) per i dati educativi.

La HIPAA acronimo di Health Insurance Portability and Accountability Act, è una legge federale statunitense promulgata nel 1996 per migliorare la portabilità e la responsabilità delle assicurazioni sanitarie. Uno degli aspetti centrali dell'HIPAA è la protezione delle informazioni sanitarie dei pazienti, impedendo la divulgazione non autorizzata di informazioni sanitarie protette, stabilendo standard nazionali per i processi di scambio elettronico di dati sanitari e contribuendo a rendere il sistema sanitario più efficiente ed efficace (ONC, 2024).

La FERPA, acronimo di Family Educational Rights and Privacy Act, è una legge federale degli Stati Uniti che protegge la privacy dei registri educativi degli studenti. Questa legge garantisce ai genitori diritti specifici riguardo ai registri educativi dei loro figli, diritti che vengono trasferiti agli studenti stessi una volta che compiono 18 anni o frequentano un istituto di istruzione superiore: questi includono l'accesso ai registri educativi, il diritto di richiedere la rettifica di registri inesatti o fuorvianti e il controllo sulla divulgazione di informazioni personali contenute nei registri senza il consenso esplicito. La FERPA si applica a tutte le istituzioni che ricevono fondi da programmi del Dipartimento dell'Educazione degli Stati Uniti, garantendo che le informazioni degli studenti siano protette e gestite in modo responsabile (U.S. Department of Education, 2024).

A livello statale il California Consumer Privacy Act del 2018 (CCPA) rappresenta un tentativo di offrire una protezione simile a quella del GDPR, e conferisce ai consumatori un maggiore controllo sulle informazioni personali raccolte dalle aziende e le sue regolamentazioni forniscono indicazioni su come implementare la legge (California Department of Justice, 2024). Questa normativa storica garantisce nuovi diritti di privacy per i consumatori della California, tra cui:

- Il diritto di conoscere quali informazioni personali un'azienda raccoglie su di loro e come vengono utilizzate e condivise;
- Il diritto di richiedere la cancellazione delle informazioni personali raccolte (con alcune eccezioni);
- Il diritto di rinunciare alla vendita o alla condivisione delle loro informazioni personali;
- Il diritto a non subire discriminazioni per aver esercitato i propri diritti ai sensi del CCPA.

Dopo l'approvazione della Proposition 24, che ha introdotto il California Privacy Rights Act (CPRA), i consumatori hanno acquisito nuovi diritti, oltre a quelli già previsti, come:

- Il diritto di correggere informazioni personali inesatte che un'azienda possiede su di loro;
- Il diritto di limitare l'uso e la divulgazione delle informazioni personali sensibili raccolte.

L'approccio normativo statunitense è frammentato e settoriale: diverse leggi regolano la raccolta e l'uso di dati personali in specifici ambiti, come nel caso della Fair Credit Reporting Act (FCRA) che è una legge federale degli Stati Uniti adottata nel 1970 che regola la raccolta, l'uso e la distribuzione delle informazioni sui consumatori relative ai report di credito. Questa legislazione mira a garantire l'accuratezza, l'equità e la privacy delle informazioni contenute nei file delle agenzie di reporting dei consumatori. Il FCRA stabilisce norme specifiche su come le informazioni sui consumatori possono essere raccolte e utilizzate, e offre ai consumatori il diritto di accedere e contestare le informazioni inesatte nei loro report di credito. La legge impone anche requisiti rigorosi per le entità che forniscono informazioni alle agenzie di credito, promuovendo trasparenza e responsabilità. Questa legge regola i data broker che operano nel settore del credito, ma non si applica a coloro che raccolgono dati per scopi non coperti da tale legge, come il marketing. Altre leggi, come l'Equal Credit Opportunity Act (ECOA) e il Fair Housing Act (FHA), influenzano indirettamente l'uso dei dati personali, vietando la discriminazione in settori come il credito e l'accesso agli alloggi.

La Federal Trade Commission (FTC) è l'agenzia che più si avvicina a un'autorità generale per la protezione dei dati negli Stati Uniti, ma i suoi poteri sono limitati. La FTC può intervenire contro aziende per pratiche ingannevoli o negligenti in materia di privacy e sicurezza dei dati, ma ha meno strumenti per affrontare pratiche ingiuste che non causano danni economici diretti

ai consumatori. Nonostante la FTC abbia più volte sollecitato il Congresso a dotarla di maggiori poteri per affrontare le problematiche legate ai data broker, tali richieste non sono state accolte.

Diversi tentativi di introdurre una legge federale che regolamenti l'uso commerciale dei dati personali, come la proposta di Obama per una "Consumer Privacy Bill of Rights", sono stati ostacolati sia dall'industria tecnologica che dai legislatori. La crescente influenza delle lobby tecnologiche e l'ampia accettazione delle tecnologie digitali rendono improbabile l'approvazione di una normativa di questo tipo nel prossimo futuro. Negli ultimi anni, i legislatori hanno iniziato a concentrarsi sui data broker specializzati in marketing, rilevando la mancanza di trasparenza nel settore. Indagini, come quelle condotte dal GAO e dal Senato degli Stati Uniti, hanno evidenziato una "scarsa conoscenza pubblica" sulle pratiche dei data broker, mentre la FCRA regola i broker che gestiscono dati per credito e impiego, quelli che si occupano di marketing sono meno regolamentati. Nonostante le indagini, non sono stati ancora evidenziati danni concreti ai consumatori, a parte un generale calo della privacy (Rieke et al., 2016).

3.6 Limiti normativi e soluzioni per la tutela della Privacy

Nonostante la presenza di normative sulla privacy come il GDPR, l'attività dei data broker è spesso caratterizzata da una scarsa trasparenza, che ostacola il controllo da parte degli individui sui propri dati personali. Questa mancanza di trasparenza ha diverse cause e rappresenta un ostacolo significativo alla salvaguardia della privacy nell'era digitale. I data broker raccolgono informazioni da una vasta gamma di fonti, tra cui siti web, app, registri pubblici e transazioni commerciali, per poi aggregarle e analizzarle, trasformandole in prodotti da vendere a terzi. Le complesse catene di approvvigionamento dei dati che ne derivano rendono difficile per gli utenti risalire all'origine delle informazioni e capire come vengano trattate, e, di conseguenza, diventa complicato per gli individui esercitare i diritti previsti dal GDPR, come il diritto di accesso, rettifica e cancellazione dei propri dati. Il GDPR richiede che il trattamento dei dati personali si basi su precise basi legali, come il consenso informato dell'individuo o la necessità di adempiere a un obbligo contrattuale; tuttavia, i data broker spesso non chiariscono in modo esplicito le basi giuridiche su cui fondano le loro operazioni, generando così dubbi sulla legittimità del trattamento dei dati e rendendo difficile per gli individui valutare la conformità alla legge. Sebbene il GDPR preveda che le aziende forniscano agli utenti informazioni chiare e concise sulle modalità di trattamento dei dati, i data broker tendono a fornire informative lunghe, complesse e piene di termini tecnici, ostacolando la comprensione da parte degli interessati. Spesso non viene comunicata con trasparenza l'identità dei terzi a cui i dati vengono

venduti, limitando ulteriormente la capacità degli individui di esercitare i propri diritti. Anche quando gli utenti sono consapevoli delle pratiche dei data broker, possono incontrare difficoltà nell'esercitare i propri diritti a causa della mancanza di procedure chiare e accessibili, e la complessità dei flussi dei dati rende spesso difficile per i data broker identificare e cancellare i dati in modo completo ed efficace. L'attività dei data broker è inoltre soggetta a una supervisione limitata da parte delle autorità competenti in materia di protezione dei dati poiché la complessità del settore e la mancanza di trasparenza rendono difficile per le autorità monitorare adeguatamente queste operazioni, favorendo così il rischio di pratiche abusive e violazioni dei diritti degli individui (Ruschemeier, 2024).

Quando i data broker e le aziende online descrivono i loro servizi, affermano spesso di trattare dati "anonimizzati" o "de-identificati". Questo significa che utilizzano tecniche come l'hashing crittografico (ad esempio MD5) per convertire informazioni come e-mail e numeri di telefono in codici unici. Tuttavia, questo processo non garantisce l'anonimato reale perché i codici risultanti continuano a fare riferimento a individui specifici e vengono chiamati pseudonimi. Grazie a metodi deterministici, le aziende possono facilmente collegare i profili digitali degli utenti, rendendo inefficaci le pretese di anonimizzazione, e, anche se non conoscono direttamente il nome di una persona, gli identificatori rimangono comunque legati alle sue ricerche e alle azioni online, facilitando il tracciamento.

Ad esempio, anche se un'azienda non sa altro su una persona oltre al fatto che ha effettuato una determinata ricerca online, quell'identificatore univoco resterà sempre collegato a quella ricerca attraverso vari database. Ogni volta che l'utente interagisce con un servizio digitale, l'identificatore rimane associato alle sue precedenti attività online e di conseguenza, anche se gli indirizzi email e i numeri di telefono sono "hashati" o "crittografati", le aziende possono comunque riconoscere i consumatori quando utilizzano altri servizi collegati agli stessi identificatori. Lo studioso di marketing Joseph Turow evidenzia che, quando un'azienda è in grado di seguire e interagire con te nell'ambiente digitale, compresi la televisione e il telefono cellulare, la sua affermazione di anonimato è priva di significato. Ciò è confermato dal fatto che le aziende spesso aggiungono informazioni offline ai dati online, rimuovendo solo nome e indirizzo per sostenere che i dati siano "anonimi". Frederik Borgesius, esperto di privacy, aggiunge che il nome è "solo uno degli identificatori legati ai dati di una persona, e non è nemmeno il più utile" nell'attuale economia del tracciamento online. Molti studiosi, regolatori della privacy e rappresentanti del settore concordano su questa visione: se un identificatore unico collega costantemente i profili digitali alla stessa persona in tutto l'ecosistema dei dati,

questi profili devono essere considerati dati personali, indipendentemente dal fatto che gli identificatori siano memorizzati in forma di testo chiaro o criptati.

Quest'opacità che caratterizza il settore dei data broker rappresenta un ostacolo significativo alla protezione della privacy nell'era digitale. Per garantire che l'innovazione tecnologica rispetti i diritti fondamentali degli individui, è essenziale un impegno maggiore da parte dei legislatori, delle autorità di protezione dei dati e degli stessi data broker, con l'obiettivo di promuovere una maggiore chiarezza nelle loro operazioni. Un primo passo potrebbe consistere nel rendere più trasparente l'intera catena di approvvigionamento dei dati, fornendo informazioni chiare sull'origine dei dati raccolti e sulle modalità del loro trattamento. Inoltre, i data broker dovrebbero spiegare in modo trasparente le basi giuridiche che giustificano il trattamento dei dati personali, garantendo così una comprensione approfondita da parte degli utenti. A questo si aggiunge la necessità di rendere le informative sulla privacy più chiare e concise, affinché gli utenti possano accedere facilmente a informazioni complete e comprensibili sulle modalità di utilizzo dei loro dati. Le procedure per l'esercizio dei diritti da parte degli utenti dovrebbero essere semplificate, in modo che possano richiedere informazioni e far valere i propri diritti in maniera semplice ed efficace, e a tale scopo, le autorità di protezione dei dati devono intensificare la supervisione e il controllo delle attività dei data broker, intervenendo con sanzioni rigorose in caso di pratiche abusive o violazioni dei diritti. Un ulteriore miglioramento potrebbe derivare dallo sviluppo e dall'adozione di tecnologie orientate alla protezione della privacy, come la crittografia e la pseudonimizzazione, che offrirebbero agli individui un maggiore controllo sui propri dati personali (Christl, 2017).

A livello pratico, una possibile soluzione per restituire agli utenti il controllo dei propri dati che oggi sono nelle mani delle piattaforme centralizzate potrebbe essere il Decentralized Social Networking Protocol (DSNP). Questo protocollo, aperto e gratuito, è progettato per decentralizzare i social media, consentendo agli utenti di mantenere il controllo completo dei loro dati e delle loro identità digitali attraverso l'uso della tecnologia blockchain.. In questo modo, i dati non sono più una risorsa aziendale, ma diventano proprietà dell'utente all'interno di un'infrastruttura pubblica gestita dalla comunità. Il DSNP consente inoltre agli sviluppatori di realizzare tecnologie che abbiano un impatto positivo e risolvano problemi concreti, senza dover sottostare agli interessi dei grandi colossi tecnologici, permettendo così un ritorno alla visione originaria del web: un'infrastruttura orientata al bene comune, in cui nessuna entità può controllare l'accesso o imporre restrizioni. Utilizzando i migliori aspetti della blockchain, con

la possibilità di scalare, il DSNP permette ai creatori di sviluppare soluzioni innovative in un ambiente che favorisce tutti (Larkin, 2023).

Grazie a questo protocollo aperto e alla tecnologia blockchain, gli sviluppatori possono riprendere il controllo delle loro applicazioni e contribuire a un processo di sviluppo collaborativo tramite il codice open source.

Conclusioni

Durante lo sviluppo di questo elaborato, è emerso con chiarezza il ruolo cruciale che i data broker svolgono nell'odierna economia digitale, quello di offrire alle aziende strumenti potenti per prendere decisioni più informate e per migliorare il targeting dei consumatori.

Attraverso l'analisi approfondita dei dati, le aziende sono in grado di comprendere meglio le esigenze e i comportamenti dei propri clienti, ottimizzando le strategie di marketing e offrendo prodotti e servizi più mirati. Questo porta non solo a una maggiore efficienza operativa, ma anche a un'esperienza personalizzata per i consumatori, che possono ricevere offerte più pertinenti e utili.

Tuttavia, questa potente capacità di analisi e personalizzazione non è esente da rischi; le sfide legate alla protezione della privacy e alla trasparenza nei confronti dei consumatori restano centrali nel dibattito sull'uso dei dati. Nonostante le normative come il GDPR e il CCPA siano state introdotte per regolare l'uso dei dati personali, il settore dei data broker continua a operare in un contesto caratterizzato da una mancanza di trasparenza e da asimmetrie informative, spesso senza che gli individui siano consapevoli della portata e dell'utilizzo dei loro dati. Le preoccupazioni relative alla sicurezza dei dati, in particolare i rischi di data breach, rappresentano un ulteriore elemento di vulnerabilità. I casi di violazioni di sicurezza, come quello di Equifax, hanno dimostrato quanto sia urgente migliorare le infrastrutture di protezione dei dati, al fine di prevenire accessi non autorizzati e di preservare la fiducia dei consumatori.

Nel corso della trattazione è emersa anche l'importanza dell'impatto sociale dell'attività dei data broker. La profilazione dettagliata dei consumatori potrebbe contribuire a creare o accentuare disuguaglianze sociali, escludendo determinati gruppi di individui da opportunità o risorse in base a informazioni che potrebbero non essere sempre accurate o contestualizzate. Un'attenzione particolare deve essere posta su questo aspetto per evitare forme di discriminazione algoritmica.

Le regolamentazioni attuali offrono una base per tutelare i diritti dei consumatori, ma sono solo un primo passo, la rapida evoluzione della tecnologia e la crescente complessità dei meccanismi di raccolta e analisi dei dati richiedono aggiornamenti costanti nelle normative e una maggiore collaborazione tra governi, aziende e consumatori per garantire che i benefici dell'economia dei dati possano essere sfruttati senza compromettere la privacy individuale.

In conclusione, per cogliere appieno i benefici dell'attività dei data broker, è essenziale che il loro operato sia guidato da principi etici e da un quadro normativo solido. Solo attraverso un

approccio equilibrato sarà possibile sfruttare le potenzialità dei Big Data mantenendo al contempo la fiducia dei consumatori e la sicurezza dei dati personali.

Guardando al futuro, il settore dei data broker continuerà a evolversi con l'adozione di nuove tecnologie come l'intelligenza artificiale e la blockchain. È cruciale che le normative si adattino costantemente a queste innovazioni, al fine di garantire una protezione adeguata dei diritti degli individui. Allo stesso tempo, un aspetto fondamentale per il futuro sarà l'educazione dei consumatori, affinché siano pienamente consapevoli di come i loro dati vengano raccolti, trattati e utilizzati, consentendo loro di esercitare un maggiore controllo sulle proprie informazioni personali. Una maggiore comprensione dei rischi e dei benefici associati all'uso dei dati personali potrà favorire richieste di trasparenza e responsabilità da parte delle aziende, spingendo verso un mercato più equo e regolamentato. Poiché il fenomeno dei data broker è di portata globale, diventa sempre più evidente la necessità di una cooperazione internazionale tra governi e aziende, per stabilire regole chiare e condivise. Soltanto attraverso sforzi congiunti sarà possibile affrontare le sfide poste dall'economia dei dati in modo efficace e garantire che i diritti degli individui siano rispettati in ogni parte del mondo.

Bibliografia

- Justin Sherman, "Data Brokers and Sensitive Data on U.S. Individuals: Threats to American Civil Rights, National Security, and Democracy", Duke University's Technology Policy Lab, Duke University, 2021.
- Wolfie Christl, "Corporate Surveillance in Everyday Life", Cracked Labs, Vienna, 2017.
- Michael Arrington, "Schmidt: Every Two Days We Create As Much Information As We Did Up To 2003", TechCrunch, https://techcrunch.com/2010/08/04/schmidt-data/, 2010.
- Oracle, 2024, "What Is Big Data?", https://www.oracle.com/big-data/what-is-big-data/.
- Excelsior, 2024, "Big Data Explained: The 5V's of Data", Medium, https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1.
- Henrik Twetman, Gundars Bergmanis-Korats, "Data Brokers and Security", NATO Strategic Communications Centre of Excellence, 2020.
- Alessandro Piva, 2024, " Le 5V dei Big Data: dal Volume al Valore", https://blog.osservatori.net/it_it/le-5v-dei-big-data.
- Kaspersky, 2024, "How to Stop Data Brokers from Selling Your Personal Information", https://www.kaspersky.it/resource-center/preemptive-safety/how-to-stop-data-brokers-from-selling-your-personal-information.
- WebFX, 2024, "What Are Data Brokers and What Is Your Data Worth?", https://www.webfx.com/blog/internet/what-are-data-brokers-and-what-is-your-data-worth-infographic/.
- Tom Kemp, 2023, "How Data Brokers Profile, Segment, and Score Us", Medium, https://tomkemp00.medium.com/how-data-brokers-profile-segment-and-score-us-5144af9a465.
- History Tools, 2024, "What Are Data Brokers and Are They Evil?", HistoryTools, https://www.historytools.org/docs/what-are-data-brokers-and-are-they-evil.
- Tibor Moes, 2023, "What is a Data Broker?", SoftwareLab, https://softwarelab.org/blog/what-is-a-data-broker/.
- Australian Competition and Consumer Commission, "Digital Platform Services Inquiry – March 2024 report on data brokers Issues Paper", Canberra, 2023.

- Tom Kemp, 2022, "A Look at the Different Types of Data Brokers", Medium, https://tomkemp00.medium.com/a-look-at-the-different-types-of-data-brokers-5ed1d360d3ea.
- Monica McCormack, 2023, "Health Data Brokers Sell Lists of Depression & Anxiety Sufferers", Compliancy Group, https://compliancy-group.com/health-data-brokers-sell-lists-of-depression-anxiety-sufferers/.
- Jasdev Dhaliwal, 2024, "What is a Data Broker?", McAfee, https://www.mcafee.com/blogs/tips-tricks/what-is-a-data-broker/.
- Acxiom, 2024, https://www.acxiom.com/customer-data/.
- Experian, 2024, https://www.experian.com/corporate/about-experian.
- Twingate Team, 2024, "Experian Data Breach", Twingate, https://www.twingate.com/blog/tips/experian-data-breach.
- Federal Trade Commission, 2024, "Equifax Data Breach Settlement", https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement.
- Corelogic, 2024, https://www.corelogic.com/why-corelogic/.
- Corelogic, 2024, https://www.corelogic.com/360-property-data/discovery-platform/.
- Corelogic, 2024, https://www.corelogic.com/real-estate/housing-trends/.
- Larry Dignan , 2014, "Oracle Acquires BlueKai, Rounds Out Its Marketing Cloud", ZDNet, https://www.zdnet.com/article/oracle-acquires-bluekai-rounds-out-its-marketing-cloud/.
- Oracle , 2024, https://www.oracle.com/corporate/.
- Justin Ménard, 2023, "The Evolution of Oracle Corporation", ListedTech, https://listedtech.com/blog/the-evolution-of-oracle-corporation/.
- Saliha, 2023, "What is LexisNexis?", Legal Inquirer, https://legalinquirer.com/what-is-lexisnexis/.
- LexisNexis, 2023, https://www.lexisnexis.com/en-us/about-us/about-us.page.
- Matthew Kosinski, 2023, "Data Breach", IBM, https://www.ibm.com/topics/data-breach.
- Megan Leonhardt, 2019, "What You Need to Know: Equifax Data Breach $700 Million Settlement", CNBC, https://www.cnbc.com/2019/07/22/what-you-need-to-know-equifax-data-breach-700-million-settlement.html.
- Federal Trade Commission , 2019, "Equifax to Pay $575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach", https://www.ftc.gov/news-

events/news/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related-2017-data-breach.

- Sara Della Piazza, 2021, "Il Caso Cambridge Analytica", DirittoConsenso, https://www.dirittoconsenso.it/2021/12/21/il-caso-cambridge-analytica/.
- Chad M.S. Steel, "Stolen Identity Valuation and Market Evolution on the Dark Web", International Journal of Cyber Criminology, George Mason University, Fairfax, 2019.
- Aliza Vigderman e Gabe Turner, 2023, "Data Broker: How to Remove Your Information", Security.org, https://www.security.org/data-removal/data-broker/.
- Reviglio Urbano (2022) : The untamed and discreet role of data brokers in surveillance capitalism: A transnational and interdisciplinary overview, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 11, Iss. 3, pp. 1-27.
- Cloudflare, 2024, "Che cos'è il Data Scraping?", https://www.cloudflare.com/it-it/learning/bots/what-is-data-scraping/.
- Glasgow Jennifer Barrett, 2018. “Data brokers: Should they be reviled or revered?” In Cambridge University Press eBooks (pp. 25–46).
- Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability", Washington, D.C., 2014.
- Esma Aïmeur, Gilles Brassard, Muxue Guo, "How Data Brokers Endanger Privacy", Transactions on Data Privacy, Université de Montréal, Montréal, 2022.
- Mengxiao Zhang, Fernando Beltrán, e Jiamou Liu, "A Survey of Data Pricing for Data Marketplaces", Journal of LaTeX Class Files, Vol. 14, No. 8, Università di Electronic Science and Technology of China, Chengdu, 2023.
- Aaron Katz, Boris Evelson, Michele Goetz, 2023, "Data Into Dollars: Can You Turn Your Data Into Revenue?", Forrester, https://www.forrester.com/blogs/data-into-dollars-can-you-turn-your-data-into-revenue/.
- Levi Kaplan, Alan Mislove, Piotr Sapieżyński, "Measuring Biases in a Data Broker's Coverage", PrivacyCon 2022, Northeastern University, Boston, 2022.
- Michael Fertik, 2013, "Rich See Different Internet Than the Poor", Scientific American, https://www.scientificamerican.com/article/rich-see-different-internet-than-the-poor/.
- Valerie L. Sizelove, 2023, "What Are Information Brokers? Everything You Need to Know", FairyGodBoss, https://fairygodboss.com/career-topics/information-brokers.

- DataIntelo, 2024, *"Data Broker Market Report"*, https://dataintelo.com/report/data-broker-market.
- Matthew Kosinski, Amber Forrest, 2023, "Data Privacy", IBM, https://www.ibm.com/it-it/topics/data-privacy.
- Xinluan Tian, Lina Chen, Xiaojuan Zhang, "The Role of Privacy Fatigue in Privacy Paradox: A PSM and Heterogeneity Analysis", MDPI, Basel, 2022.
- Dobrica Savić, "Information Fatigue Syndrome and Digital Burnout", The Grey Journal (TGJ), 2023.
- Nina Gerber, Paul Gerber, Melanie Volkamer, "Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior", Elsevier, Amsterdam, 2018.
- Commissione Europea, 2024, “Legge sui dati”, https://digital-strategy.ec.europa.eu/it/policies/data-act
- Commissione Europea, 2024, “Data Act – Questions and Answers”, https://ec.europa.eu/commission/presscorner/detail/en/qanda_22_1114
- Unione Europea, "Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali e alla libera circolazione di tali dati (Regolamento generale sulla protezione dei dati)", Gazzetta ufficiale dell'Unione Europea, L 119/1, 2016.
- Aaron Rieke, Harlan Yu, David Robinson, and Joris von Hoboken, "Data Brokers in an Open Society", Upturn, prepared for the Open Society Foundations, London, 2016.
- California Department of Justice, "California Consumer Privacy Act (CCPA)", https://oag.ca.gov/privacy/ccpa, 2024.
- Office of the National Coordinator for Health Information Technology (ONC), "HIPAA Basics", U.S. Department of Health & Human Services, https://www.healthit.gov/topic/privacy-security-and-hipaa/hipaa-basics, 2024.
- U.S. Department of Education, "What is FERPA?", U.S. Department of Education, https://studentprivacy.ed.gov/faq/what-ferpa, 2024.
- Hannah Ruschemeier, 2024, "Data Brokers and the Limits of the GDPR", Verfassungsblog, https://verfassungsblog.de/datatrade-eu-gdpr-privacy/.

– Martina Larkin, 2023, "We Need to Break Big Tech's Data Broker Industry", TechCrunch, https://techcrunch.com/2023/10/02/we-need-to-break-big-techs-data-broker-industry/.